

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/2401>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

A Study of Methods of Evaluating Multimedia Materials for Language Learning

Saima Nawaz Sherazi

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in
Applied Linguistics & English Language Teaching

Centre for English Language Teacher Education

University of Warwick

January 2007

**THESIS
CONTAINS
CD**

TABLE OF CONTENTS

Table of Contents ii

List of Tables and Figures.....vi

Abbreviations..... viii

Acknowledgementsx

Declaration.....xi

Abstract.....xii

1 Chapter One: Background to the research 1

1.1 Introduction 1

1.2 CALL-MM Materials: an overview of development..... 2

1.2.1 Principles underlying the use of CALL materials 5

1.2.2 Practical activities offered in CALL materials 6

1.2.3 Essential Academic Skills in English (EASE) 11

1.3 The significance of this research 14

1.3.1 The study 16

1.4 The structure of this thesis 17

2 Chapter Two: A Framework for the Study 20

2.1 Introduction 20

2.2 Evaluation in education 21

2.2.1 The relationship between Research and Evaluation 21

2.2.2 Models and approaches in educational evaluation 22

2.2.3 Research strategies and techniques in educational evaluation 27

2.3 Evaluation of language teaching materials 27

2.3.1 Approaches to materials evaluation..... 28

2.3.2 ELT materials evaluation methods..... 30

2.3.3 Concluding remarks..... 34

2.4 Evaluation methods in Human Computer Interaction 34

2.4.1 Usability evaluation..... 35

2.4.2 Concluding remarks..... 42

2.5 Evaluation in CALL 42

2.5.1 The nature of Hubbard’s work..... 43

2.5.2 CALL evaluation studies and their methods 48

2.5.3 Concluding remarks..... 58

2.6 A synthesis..... 59

2.7 A selection for closer study 61

2.7.1 Focus groups..... 62

2.7.2 Retrospective protocols 65

2.7.3 Questionnaires..... 68

2.7.4 Activity monitoring and keystroke logging 71

2.8 Summary 72

3 Chapter Three: The Conduct of the Study..... 74

3.1 Introduction 74

3.2 The research question 74

3.3 Research paradigm, strategy and approach 75

3.4 The research design 75

ii

3.5	The pilot study.....	76
3.5.1	Methods for trial	76
3.5.2	The procedure	78
3.5.3	Findings and discussion.....	79
3.6	Secondary research questions	82
3.6.1	Rationale for the research questions	83
3.7	The main study.....	84
3.7.1	Participants.....	84
3.7.2	The procedures of the study.....	85
3.7.3	Focus groups.....	88
3.7.4	Retrospective protocols.....	92
3.7.5	PLUM and SUMI questionnaires	96
3.7.6	Activity monitoring and key stroke logging.....	101
3.8	Criteria for evaluating methods for multimedia CALL materials.....	103
3.7.7	Cost effectiveness	104
3.7.8	Ease of use	104
3.7.9	Bias: Researcher or Respondent.....	105
3.7.10	Ecological validity /Intrusiveness or Interference.....	105
3.7.11	Nature of information.....	105
3.7.12	Immediacy of response	106
3.7.13	Usability measure	107
3.7.14	Robustness of method	108
3.8	Summary.....	109
4	Chapter Four: Focus Groups as an Evaluation Method	110
4.1	Introduction	110
4.2	Focus group findings and analysis	110
4.2.1	Initial focusing exercise.....	110
4.2.2	Main focus group discussion	112
4.3	Focus group as an evaluation method	129
4.3.1	Two focus groups	130
4.3.2	Role of the moderator.....	130
4.3.3	Salient features of focus group evaluation	131
4.3.4	The method's performance on the criteria of evaluation.....	138
4.4	Summary.....	142
5	Chapter Five: Retrospective Protocols as an Evaluation method.....	143
5.1	Introduction	143
5.2	Findings and analysis	143
5.2.1	Evaluative comments from oral protocols	143
5.2.2	Evaluative comments from e-mails.....	157
5.3	Main findings of the evaluation.....	161
5.4	Retrospective protocols as evaluation method.....	162
5.4.1	The method's performance on the criteria of evaluation.....	166
5.4.2	Comparison of the two variations of retrospective protocols	169
5.5	Summary.....	170
6	Chapter Six: Questionnaires as an Evaluation Method	171
6.1	Introduction	171
6.2	PLUM questionnaire	171
6.2.1	Pre-use questionnaire	171

6.2.2	PLUM post use questionnaire	173
6.2.3	Analysis of Listening to Lectures.....	173
6.2.4	Analysis of Seminar Skills	181
6.3	SUMI: Software Usability Measurement Inventory	187
6.3.1	Interpreting SUMI	187
6.3.2	SUMI results for Listening to lectures.....	191
6.3.3	SUMI results for Seminar Skills	200
6.4	Questionnaires as evaluation instruments	207
6.4.1	Overall results from the perspective of an evaluation of EASE.....	208
6.5	PLUM and SUMI's performance on the evaluation criteria.....	209
6.6	Summary	212
7	Chapter Seven: Activity Monitoring as an Evaluation Method	213
7.1	Introduction	213
7.2	Data reduction and analysis procedures	213
7.2.1	Navigation patterns	214
7.2.2	Time usage	215
7.2.3	Narratives of navigation	217
7.3	EASE <i>Listening to Lectures</i>	217
7.3.1	Flowchart of Listening to Lectures Unit 2.....	217
7.3.2	Flowchart of Listening to Lectures Unit 6.....	220
7.3.3	EASE Listening to Lectures Unit 2: Structures and Organizations.....	222
7.3.4	EASE Listening to Lectures Unit 6: Argumentation.....	228
7.3.5	Five Students navigation through Unit 6 of EASE Listening to Lectures	229
7.4	EASE <i>Seminar Skills 1: Presentations</i>	231
7.4.1	Flowchart of Seminar Skills Unit 1	232
7.4.2	Flowchart of Seminar Skills Unit5.....	234
7.4.3	Description of EASE Seminar Skills.....	235
7.4.4	Five Students navigation through Unit 1 of Seminar Skills	238
7.4.5	EASE Seminar Skills Unit 5:.....	239
7.4.6	Five students' navigation through Unit 5 of Seminar Skills.....	241
7.5	Interpretations and findings from two perspectives	242
7.5.1	From the perspective of the evaluation of EASE	242
7.5.2	Activity Monitoring as an evaluation method	247
7.6	The method's performance on the criteria of evaluation.....	247
7.7	Summary	251
8	Chapter Eight: Discussions of Findings and Conclusion.....	252
8.1	Introduction	252
8.2	An overview of the study	252
8.3	Revisiting the research questions	253
8.3.1	Research Question 1.1	253
8.3.2	Research Question 1.2	255
8.3.3	Research Question 1.3	256
8.3.4	Research Question 1.4	257
8.4	Towards best practice in MM materials evaluation	259
8.4.1	Comparative Performance of methods on the Evaluation Criteria	260
8.4.2	Different combinations of the model	264
8.5	Contribution.....	266
8.6	Suggestions for Further Research.....	267
8.7	Concluding remarks.....	270

Bibliography	271
Appendices	287
Appendix 1 .1: Description and Appearance of Materials.....	288
Appendix 2.1 EASE Listening to Lectures CD-ROM	299
Appendix 3.1 EASE Seminar Skills CD-ROM	300
Appendix 4.1 Focusing exercise	301
Appendix 4.2 Points From FG Discussion.....	301
Appendix 4.3 Long Table.....	302
Appendix 4.4 Focus Group Colour coded Sample.....	306
Appendix 4.5 Nvivo Coding Report of Focus Group 2.....	310
Appendix 4.6 Selected Comments from focus groups	312
Appendix 4.7 Comments from section 4.4.3	318
Appendix 5.1 Comments in full from Section 5.2.1	321
Appendix 5.2 Comments in full from Section 5.2.1.2.....	330
Appendix 5.3 Comments in full from Section 5.2.1.3.....	332
Appendix 5.4 Comments in full from Section 5.2.2.....	335
Appendix 6.1 PLUM pre program use questionnaire	343
Appendix 6.2 PLUM post program use questionnaire	343
Appendix 6.3 Reliability test <i>Listening to Lectures</i>	344
Appendix 6.4 Reliability test <i>Seminar Skills</i>	345
Appendix 6.5 SUMI Questionnaire	346
Appendix 6.6 Sumisco Report	348
Appendix 7.1 Key logger reports	357
Appendix 7.2 Description of Units.....	359
Appendix 7.3 Students' Navigation Accounts	364
Appendix 7.4 Time usage worksheet.....	370
Appendix 7.5 CD containing additional materials:visio flowcharts.....	371

LIST OF TABLES AND FIGURES

List of Tables

Table 1.1 CALL Exercise types9

Table 2.1 McGrath's Checklist (2002) 31

Table 2.2 Evaluation Techniques in HCI Usability Evaluation 39

Table 2.3 Methods across Disciplines: A Composite List..... 59

Table: 3.1 Results of the Pilot Study..... 81

Table 3.2 Focus Groups Participant profile 89

Table 4.1 Focussing exercise statements..... 111

Table: 4.2 Results of the ranking exercise..... 111

Table 4.3 Discussion questions for focus groups..... 112

Table 4.4: Summary of emergent points 114

Table 4.5 Areas of agreement and disagreement..... 115

Table 4.6: Biases in Learner Response 116

Table 4.7: Negative/ Positive Results..... 117

Table 4.8 Results..... 117

Table 4.9: Themes and Categories 118

Table 4.10 Performance on Criteria..... 139

Table 5.1: Themes and Categories 144

Table 5.2 Positive and Negative Comments 146

Table 5.3: ROP and RRWA Performance Matrix..... 166

Table 6.1 PLUM Question 2 *Listening to Lectures* 180

Table 6.2 PLUM Question 2 *Seminar Skills*..... 186

Table 6.3 Comparison of Day 1 and Day 2 Results for *Listening to Lectures* 199

Table 6.4Comparison of Day 1 and Day 2 SUMI results of *Seminar Skills* 206

Table 6.5 PLUM and SUMI' s Correspondence in Scales..... 208

Table 6.6 Questionnaires' Performance on Criteria..... 209

Table 7.1 Key logger reports from two students..... 215

Table 7.2 Questionnaire's performance on Criteria..... 248

Table 8.1: Comparative Performance of Methods..... 262

Table 8.2: Combinations of Methods..... 265

List of Figures

Figure: 1.1 A screenshot of EASE Vol: One <i>Listening to Lectures</i>	12
Figure 1.2: A Screenshot of EASE Vol: Two <i>Seminar Skills</i>	13
Figure 2.1 Evaluation Methods in Research (Bennet, 2003:58-59).....	36
Figure 2.2 Usability Evaluation Methods in HCI (Whitefield <i>et al.</i> 1991)	36
Figure 2.3 Usability Evaluation Method Classes (Ivory & Hearst 2001:473)	37
Figure 2.4 HCI Evaluation Techniques Using Experts and End-users	37
Figure 2.5 Hubbard's Development Module (1996:19)	45
Figure: 2.6 A Venn diagram illustrating the positioning of this study	73
Figure 6.1 PLUM Question 2 <i>Listening to Lectures</i>	181
Figure 6.2 PLUM Question 2 <i>Seminar Skills</i>	187
Figure 6.3 SUMI Results of <i>Listening to Lectures</i>	193
Figure 6.4 DAY-1 SUMI results of <i>Listening to Lectures</i>	199
Figure 6.5 SUMI results of <i>Seminar Skills</i>	200
Figure 6.6 DAY-1 SUMI results of <i>Seminar Skills</i>	206
Figure 7.1 Generic Flowchart of EASE Vol 1: Listening to Lectures Unit 2.....	218
Figure 7.2 Closer look at start page of unit 2 Vol1	219
Figure 7.3 Generic Flowchart for Unit 6 of Listening to Lectures	220
Figure 7.4 Closer look at Unit 6 Vol 1	221
Figure 7.5: PC53 (session 1 on Unit2) Navigation Pattern shown by red coloured lines.....	223
Figure 7.6 Student PC53 consulting the Grammar book	225
Figure 7.7 Navigation pattern Unit 2.....	226
Figure 7.8 Time Spread Unit 2 Listening to Lectures.....	226
Figure 7.9 Flowchart of one student's navigation PC 03.....	229
Figure 7.10 Unit 6 Vol1 Navigation picture	230
Figure 7.11 Unit 6 Vol 1 Time Usage.....	230
Figure 7.12 EASE 2 Unit 1 Flowcharts	232
Figure 7.13 <i>Seminar Skills</i> Unit1: Closer Picture.....	233
Figure 7.14 <i>Seminar Skills</i> Unit 5 Flowchart.....	234
Figure 7.15 Closer Look at <i>Seminar Skills</i> Unit 5	235
Figure 7.16 PC30 working on Unit 1 <i>Seminar Skills</i>	237
Figure 7.17 Navigation picture Unit 1 <i>Seminar Skills</i>	238
Figure 7.18 Time Spread graph of Unit 1 <i>Seminar Skills</i>	238
Figure 7.19 Student PC31 working on Unit 5 Seminar Skills	240
Figure 7.20 Navigation Picture Unit 5 <i>Seminar Skills</i>	241
Figure 7.21 Time Usage <i>Seminar Skills</i> Unit 5.....	241
Figure 7.22 PC13 Checking cricket scores. Check clock 18.06.31	244
Figure 7.23 Pc13 Rewarding with cricket scores. Check clock 18.06.41 (10 sec interval)	245
Figure 7.24 Screenshot showing PC13's note-taking effort	245
Figure 7.25 : A student multitasking. Time 19:12:46	246
Figure 7.26 Accessing databases. Time: 19:12:56	247
Figure 8.1: The positioning and contribution of this study.....	266

ABBREVIATIONS

Act Mon: Activity Monitoring	SAT: Scholastic Aptitude Test
AMT: Activity Monitoring and Tracking	SL: Second Language
CALL: Computer Assisted Language Learning	STM: Short Term Memory
CBI: Computer Based Instruction	SUMI: Software Usability Measurement Inventory
DCT: Dual Coding Theory	TAM: Tracking and Activity Monitoring
EASE: Essential Academic Skills in English	TAP: Think Aloud Protocols
EFL English as a Foreign Language	TESOL: Teaching of English to Speakers of Other Languages
ELT: English Language Teaching	UE: Usability Evaluation
ESL: English as a Second Language	UQ: Usability Questionnaire
FE: Further Education	UWT: User Walk Through
FA/FSc: Fellow of Arts/Science	VP: Verbal Protocols
FG: Focus Groups	
HCI: Human Computer Interaction	
HE: Higher Education	
IELTS: International English Language Testing System	
IS: Interest Section	
LAN: Local Area Network	
LUMS: Lahore University of Management Sciences	
LTM: Long Term memory	
MM: Multimedia	
PLUM: Programme on Learner Use of Media	
ROP: Retrospective Oral Protocols	
RP: Retrospective Protocols	
RRWA: Retrospective Reflective Written Accounts	

For:

Ammi, Abbu & Ali Bhai

& especially for

Daniyal, Minabil, Seher & Aiza

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisors Dr. Hilary Nesi and Dr. Julia Khan for their inspiring guidance and kind support throughout the period of writing this thesis.

I owe sincere gratitude to Mr. Rod Revell, Dr. Gerard Sharpling, and the ELT Unit at CELTE, for being supportive mentors and colleagues, particularly during my pilot study in the pre sessional. A special thanks to Mr. Tim Kelly for digitizing my video-based data and sharing design and development information about EASE. I am grateful to Ms. Tilly Harrison for sharing her work on key loggers and to Dr. J. Kirakowski for assistance with statistical analysis of SUMI. My thanks are also due to Dr. Keith Richards, Dr. Richard Smith and Dr. Sultan Erdogan-Smith for their advice on my work.

I would also like to express sincere gratitude to Dr. William Crawley and the Charles Wallace Pakistan Trust for the award of a fellowship grant; to Dr. Kay Sanderson and Ms. Kathy Williams for my association with the Warwick Skills Programme; to the faculty and students of the Lahore University of Management Sciences (LUMS) particularly to Dr. Syed Zahoor Hassan, Dr. Faisal Bari and Ms. Naheed Haq a special thanks.

This effort only became possible with the love and support of my family. Sincere appreciation is due to my parents for their love, prayers and unwavering support, to my brothers Ali, Bilal and Jawad for their kindness and generosity, and to my children Seher, Daniyal and Minahil for their love and patience.

My friends provided tremendous support with their words of wisdom, encouragement and good wishes. To all my friends at CELTE, in England and in Pakistan I owe a deep debt of gratitude.

DECLARATION

I hereby declare that this thesis is entirely my original work and has not been submitted for any higher degree in another university.

The papers listed below were presented at CELTE's postgraduate conferences during the period of the preparation of this thesis. Content from these papers has been adapted for use in the thesis.

Sherazi, S.N. (2006). Evaluating Evaluation methods for CALL Multimedia Applications. CELTE Postgraduate Summer Conference June 2006. University of Warwick

Sherazi, S.N. (2004). How effective design principle of multimedia language teaching materials can be used as evaluation criteria for evaluating language teaching applications .CELTE Postgraduate Summer Conference June 2004.University of Warwick

ABSTRACT

There is a long tradition of paper based materials evaluation in ELT, but at this juncture, a scarcity of studies on ELT Multimedia (MM) materials evaluation. Such studies as have been undertaken have tended to adopt the perspective of the materials developer rather than the end user. But there have been no developed studies of evaluation methods which could be adopted by potential users. Despite calls being made for systematic evaluation, not many proposals have been developed, and there has been little exploration of potential best practice, or of the 'goodness of fit' between methods and evaluation purposes. This study aims to investigate evaluation methods in order to establish best practice in the evaluation of multimedia CALL applications, with a focus on learners' experience and opinions, and with the aim of enabling potential teacher-users of CALL materials to gauge the suitability of materials for their learners.

Chapter one of this thesis provides a rationale for this study and an overview of the background to this research. Chapter two presents a review of literature undertaken in four domains: educational evaluation and research methods; Human Computer Interaction (HCI) usability evaluation methods; ELT materials evaluation; and studies of CALL materials in use. Chapter three focuses on the design and conduct of the study by explaining how some methods of evaluation were trialled in a pilot study and four were selected for adoption and analysis in the main study. The methods selected were *Focus Groups*, *Retrospective Protocols*, *PLUM* and *SUMI Questionnaires* and *Activity Monitoring*. To determine the qualities and limitations of these methods, a set of criteria was developed from the literature on software usability evaluation methods in HCI and a broader literature on educational evaluation and research methods. The four data study chapters (4-7) each discuss one of the chosen methods and describe how the method was operationalised in an evaluation of learner responses to multimedia software. The final chapter draws together the discussion of the findings and presents different proposals for best practice.

The focus in the discussion of findings is on how the chosen methods performed according to the set of criteria. The findings confirm that focus groups and questionnaires are quick and efficient methods whereas retrospective protocols and activity monitoring provide more detailed and protracted data. Teacher evaluators can be guided by the objectives of their evaluation to explore different combinations of these methods.

Participants in the pilot were 12 ESL students from the University of Warwick and in the main study 45 Freshman/sophomore students from a university in Pakistan. The materials used to operationalise methods were the EASE CD-ROMS *Listening to Lectures* and *Seminar Skills 1: Presentations*.

The research contributes to the field in undertaking an in-depth and extensive study of evaluation methods applicable to CALL materials, which adopt a learner-centred perspective, and conform to sound principles within educational evaluation, yet which draw on practice in the field of HCI, since this expertise is so relevant in the rapid development of multimedia materials for use in ELT. Moreover, by developing the composite set of core criteria this study has created a tool which practitioners in the field can use to select most appropriate methods for their particular evaluation purposes.

CHAPTER ONE: BACKGROUND TO THE RESEARCH

1.1 Introduction

The rapid development of the Internet and computer based learning technologies has led to enormous spending by educational establishments on hardware and software. Fear of falling behind, and the desire to provide learners with the best opportunities, are at the heart of this drive to adopt innovative technological resources. Changes are rapid, and are likely to continue to be so, as the increased power of microprocessors and affordable pricing of Windows-based software makes developments widely accessible. In the field of ELT the drive to adopt and develop opportunities often leads to the purchase of new language teaching/learning software.

Within institutions, the selection of hardware and software to be purchased is a significant responsibility. Decisions about the purchase of hardware may rest with central IT service providers – (and there will be frequent challenges to be met as operating systems develop and change, and hardware needs to accommodate them). But it is more often individual departments within an HE/FE institution or a language school who need to make decisions on the purchase of software, in the same way as they might select paper based materials and audio visual materials, in order to meet the needs of students and support them in achieving the learning aims of the curriculum. Software reviews in magazines and publishers' catalogues offer help to potential purchasers trying to choose between available options, but they cannot address the specific needs or opinions of users in a particular situation. What approach to selection might best be used therefore? Sums being spent are considerable, and decisions crucial. It is clear therefore that the evaluation processes which underlie decision-making need to be as effective as possible. This is the key justification for engaging in research into these processes.

Issues other than financial can also be pertinent. For example, since teaching through electronic materials still, for some, means a paradigmatic shift from exclusive use of 'traditional' face-to-face teaching, there is in some places an internal resistance to adopting such material at all since it signals change, which in turn may be seen to signal that previous practice was inadequate. Indeed, any techno-phobic members of a department may hope for failure when it is used. In such circumstances, departments may play safe by investing in self-access electronic

materials, hoping thus to transfer the onus of success onto the learner. This may be an important factor in evaluating materials with a view to purchase, and methods of doing the evaluation may need to be taken into account.

Evaluation methods and procedures also have to be able to adjust to the wide variety of materials available, and the fact that at present, authorship is very diverse. Some materials are being produced by very experienced designers; others by novice enthusiasts of technology. The absence of any formal quality control measures in the domain of ELT Multimedia materials leaves great responsibility in the hands of potential selectors to evaluate quality and gauge learners' responses.

This study explores various approaches to evaluation in order to identify possible best practice in evaluating computer-based materials for English language teaching and learning. It does so by operationalising a number of methods and conducting an evaluation in order to study the performance of these methods. Their effectiveness is gauged in relation to a set of criteria established as part of the methodology of the study.

1.2 CALL-MM Materials: an overview of development

In English language teaching, interest in computer assisted language learning (CALL) has been gradually increasing since the 1980's. However, the magnitude of this interest has grown in the last decade or so because of increased accessibility and speed of connections, and growing sophistication of hardware and software. Beatty (2003:7) defines CALL as *"any process in which a learner uses a computer and, as a result, improves his or her language"* and suggests that this definition is intentionally broad to accommodate CALL's changing nature.

An explanation of CALL in practice is given by Felix (2003b). She gives convincing reasons for using computer technologies both as supplements to regular classrooms and for improving the quality of both traditional distance education and regular face to face teaching. Computer based learning exploits both web based/ LAN based learning and CD-ROM based learning. According to Felix (2003b) online and CD-ROM based learning can be categorized into two major forms. Firstly there are stand-alone online courses that try to function as virtual classrooms in which the technology behaves both as tutor and tool (Felix 2003b). Secondly there are add on or supplemental activities to classroom teaching or distance education courses in which technology is used primarily as a tool and communication device. The best current approaches use technology as a tool and the

objective is to create learning environments which would otherwise not be available in the classroom, in which a creative teacher can set up learning tasks that are stimulating and engaging and which take individual student differences into account (Felix 2003b).

Paivio's (1971, 1986) work on the psychology of learning and the application of *Dual Coding Theory* (DCT) to education provided a first rationale for the use of multimedia materials. DCT attempts to explain key "... mental structures and processes: the structures are associative networks of verbal and imaginal representations, and the processes concern the development and activation of those structures..." (Clark and Paivio 1991:151). DCT can provide explanations as to how learners learn through association and explains the cognitive processes that are involved in the learning of languages through multimedia materials (Ibid: 149). Multimedia materials are able to present verbal, non verbal , sound , colour, graphics, audio and video content in one software application, in any combination, carrying key attributes of producing usability and functionality of system (McKerlie and Preece 1993).

Different agencies and individuals are developing multimedia software for language teaching. A count of publishers listed on TESOL CALL IS website reveals 212 publishers publishing CALL materials. The growth of well established publishers of commercially available materials suggests that language learning software have vibrant markets all over the world. A survey of ELT MM materials shows a surge of development in the last two years. More than 300 CD-ROM s specially designed for EFL were available in 1999 (Eastment 1999). There are no equivalent statistics available for 2006, but a count of the CALL software listed on TESOL CALL-IS Software list showed approximately 744 items of commercially available software. If we consider one publisher CLARITY as an example: it alone publishes 45 titles of stand-alone and network based CD-ROMs, covering numerous areas: Teacher Development; Authoring suites; Business English; Dictionaries; Exam preparation; Grammar; IELTS preparation; Study skills; Integrated Skills; Pronunciation; Reading; Speaking and Listening; Teacher's Tools; Vocabulary; Writing and Young learners. Amongst its CDs, some are stand-alone teaching CD-ROMs (i.e electronic interactive textbooks with MM materials), and some are CD-ROMs accompanying textbooks. Most textbooks currently being published for language teaching purposes are accompanied by CD-ROMs, and stand-alone CD-based courses are also increasing day by day.

Other than commercial publishers who have teams of developers working for them there are the independent freelance developers of interactive multimedia software who publish their own work, who may use many channels for marketing: brochures, leaflets, newsletters, websites, or catalogues from for example KELTIC or publishers with extensive ELT collections such as Oxford University Press. These independent developers have cornered niche markets concerned with specific aspects of language learning. For instance some develop software for listening comprehension, others for grammar, yet others for pronunciation and so on. These areas may in their turn be broken down into narrower specialisms. For example, within the area of pronunciation, an academic produced the advanced level *Streaming Speech: listening and pronunciation for advanced learners of English* (Cauldwell 2002) which focuses on stress and intonation patterns of speech, whereas Sky Pronunciation Suite (Sky software house) focuses on the more basic level of the phonemic alphabet, similar sounds and phrasal stress.

The third kind of ELT software developers are teams of faculty and programmers in ELT departments of universities. Two such teams working in the EAP context in the UK are at CELTE, University of Warwick who have developed EASE (Kelly, Nesi and Revell, 2001; Kelly, Nesi and Richards, 2004) and the Division of Languages and Intercultural Studies at Anglia Ruskin University who developed *Excel at Academic English* (Ferney and Waller, 2000).

The challenges faced by a CALL MM software developer relate both to creative design and programming. Materials must engage learners' interest, present language content and learning tasks as effectively as possible, and consider the design of the user interface using colours, graphics and animation in a way that exploits the potential of the software to help learners achieve the learning objectives. The software also has to be compatible with all the different hardware and operating systems. According to some language teaching software developers, aesthetic design is as important as functionality as a means of supporting learning (Don Friend, personal communication, 14 August 2005; Tim Kelly, personal communication, 14 August 2005).

Learners as the end users of these products are continually evolving as they learn to work with sophisticated advances in software technology particularly the state of the art advances in gaming software and are raising the bar of expectation from learning software. Well-conceived and well-designed MM materials can also customise programs for individual learners by being able to provide a wide choice of options (Brett 2000; Jiménez and

Pérez 2002; Pérez 2000). Learning is customised by the individualised nature of feedback: “Combined with the feedback element MM is the most sophisticated self-use tool to date...” (Brett 2000). This opportunity allows learners to explore and experiment with their learning processes and preferences, giving them greater control of their learning – leading to empowerment and useful levels of awareness (Beatty 2003; Hedberg Harper and Brown 1993; Watts, 1997). Positive attitudes on the part of teachers are also very important in maximising the benefit to be gained from use of MM materials.

In the context of higher education and teaching innovation, teachers’ attitudes to Computer Based Instruction (CBI) are certainly important. One of the most significant challenges in keeping abreast of technological change is staff motivation to update and renew their course materials and innovate their delivery. There may be some tension between institutional pressures to keep up to date with e-learning innovations, and the perceived effort needed to do so. CBI/CALL on the one hand is propelled forward by the enthusiasm of motivated professionals, while on the other hand there is little guidance and extrinsic motivation for teachers using new technologies as tools.

1.2.1 Principles underlying the use of CALL materials

Multimedia materials have been defined as an integrated technology that stores and retrieves information in several media forms combining texts, graphics, animations, audio and video either on a computer disc or on the internet (Hall 1996; McKerlie and Preece 1993; Tolhurst 1995). The contents are hyperlinked and integrated, presenting the learner with options to access and navigate through a vast amount of information (Beatty 2003; Watts 1997).

Our understanding of the term *multimedia* in relation to materials is dependent on our understanding of *hypertext* and *hypermedia*. Hypertext can be defined as links (also called hotlinks or hyperlinks) between texts, which are indicated by a blue line and which, when clicked on, take the reader to the linked page that is being sought (Beatty 2003:37). This extra hyperlinked information can be seen as replacing the additional materials traditionally presented in footnotes and annotations, except that there is no length limitation on the notes. Hypermedia operates the same way as hypertext but instead of linking only to text, also links to different media such as sound, images, animations and video (Beatty 2003:39). Hence Beatty’s definition: “the term *multimedia* is

used to encompass the non linear organisation of text in *hypertext* and the non linear and multiple information formats referred to in *hypermedia*” (Beatty 2003:41).

The term Multimedia thus implies certain attributes which can be described as “multisensory”, “multimodal”, and “multichannel” (Marmolin 1991). Building on this understanding of multimedia, Hoogeveen defined multimedia as a system or object in which “... multiple perceptual representation media, such as speech, music, text, graphic, still, animation and video, are used in an integrated manner.” (1997: 151).

One potential advantage of multimedia software is that it can take into account the negotiated interaction model of language acquisition and enable the learner to work autonomously. (Jiménez and Pérez 2002; Pérez 2000). Negotiation during interaction provides access to different language forms and this aids understanding (Pica 1994). Brett (2000) argues that the process of negotiating afforded by multimedia software aids acquisition, and interactional moves such as clarifying, checking, confirming and redoing develop understanding. The language input can be provided through a video, followed or complemented by meaning-focused exercises or tasks, and immediate feedback can check understanding and aid learning. Written subtitles or transcripts of spoken texts can resolve problems related to pronunciation or pace of speech. Dictionaries and glossaries can also support understanding.

The nature of MM materials is such that they have the potential to create conditions where interactivity, learner-centredness and autonomous learning can thrive, and immediate individual feedback can become available. These features can contribute a lot to learner motivation (Crowther, Keller and Waddoups 2004; Liu, Ayersman and Reed 1995; Stepp-Greany 2002). Students may work at their own pace when they use CALL programs - a key feature of autonomous learning (Jiménez and Pérez 2002; Pérez 2000). In other ways too, the design of MM materials allows for adjustments which cater to individual learning needs and take account of the learning preferences of students (Watts 1997). These features are consistent with the principles of constructivism and communicative language teaching.

1.2.2 *Practical activities offered in CALL materials*

CALL's relationship with language learning and teaching theory has moved from being 'theory driven' in the early days to being 'theory accommodated' in order to be compatible with pedagogical theories underlying

teaching approaches such as a communicative approach or autonomous learning (Scholfield and Ypsaldis 1994:62-63).

However, CALL experts sometimes argue that developments in CALL have been led by improvements in technology rather than pedagogy. A key concern of ELT practitioners is that although the technological improvement in this field is very fast, pedagogical innovation is not keeping pace with technological innovation (Laurillard 2003: ix). Levy (1997:41) emphasises that “evaluation studies are crucial if CALL is not to be entirely technology led, and if we are to build upon prior successes.” Recent contributions to the field, by focussing on “...the reality of learner’s experience are capable of forcing the technology to serve pedagogy, not vice versa” (Laurillard 2003: ix). They argue that advances and innovation in the broad field of online technology are potentially applicable to CALL materials. But whilst materials may have moved from basic gap filling to sophisticated interactive multimedia presentation “... in terms of pedagogy, the new and improved have not always replaced the old and tired.” (Beatty 2003:11). The following section attempts to show how “this lack of concerted progress” (Beatty 2003:11) is reflected in the work of materials designers over the last three decades by focusing on the exercise types that have appeared in CALL materials.

According to Goodfellow (1995) the first issue in designing software programs for language teaching is the extent to which activities are consistent with the implicit demands of learning objectives, whether they relate to the learning of vocabulary, grammar, pronunciation or any other aspect of language. Secondly, it is the extent to which the personality, feel, flavour or worth of the interaction is generated by an exercise. This depends on the choice of diction and the way concepts are explained or the level at which they are explained, the pedagogical principles underpinning the design and the program’s adaptability to a learner’s individual needs. The third issue is whether material is suitable to the interactive and computational environment and lends itself naturally to the medium or is just an online version of the paper-based activity (Goodfellow 1995). Does it fully explore and utilise the ‘...the interactive potential generated by storage and access capabilities of the medium.’? (Goodfellow 1995: 205)

Even for paper-based language teaching materials a typology of exercise types or task types is difficult to formulate. Attempts have been made by Dodson (1967: 178-80) Grellet (1981: 12-13) Maley (1998:288-91) to name a few. Most of them have expressed difficulty in limiting the number of exercise types; “There are a

number of difficulties involved in setting up a typology...since the exercises are a mixture of activities...techniques which can be adapted to a range of activities....” (Ahmad *et al*1985:102). McGrath (2002:113) comments on the need for such a typology and expresses the myriad possibilities that each exercise generates: “As with any good typology, the number of ideas that can be generated by each option is limited only by the user’s imagination”. Such a typology would assist material designers of all contexts including CALL.

Davies and Higgins (1985:14) identified the following CALL exercise types: completion or gap filling exercises; substitution exercises; transformation exercises; recognition exercises; production exercises. Higgins and Johns (1984 :35-87), proposed a somewhat different typology, identifying question-answer dialogues, quizzes, multiple choice tests, open ended vocabulary tests, grammatical drills, gap filling exercises, generative question answer routines, jumbled letters, words or sentences, cloze tests, and games and mazes. The range of exercise types that can be exploited by CALL was further expanded by Ahmad *et al* (1985:102). A composite list of exercise types based on a review of the literature is given in Table 1.1 below.

In the table, the capitalized column headings denote umbrella terms under which the different exercises can be subsumed. ‘Text deletion’ and ‘text reconstruction’ are two kinds of *manipulation* of text (‘manipulation’ stated in small case in the fourth row just above text deletion and text reconstruction). Bold case headings within columns name the kind of exercises in that category. ‘Target Spotting” is an umbrella term which requires learners to spot particular words (or objects within pictures) which may be out of context or peculiar. This type of exercise can be exploited to spot similarities or differences or peculiarities or encouraging ‘editing and proof reading’ work.

Table 1.1 CALL Exercise types

TARGET SPOTTING	
Proof reading	Target spotting analysis
Comparison/contrast	Words in context
Word Forms	Word Search
CREATING A TEXT	
Free text	
MEDIA TRANSFER	
Dictation	Speech recognition
Manipulation	
TEXT DELETION AND MANIPULATION	TEXT DELETION AND RECONSTRUCTION
Jumbled Letters, Words, Sentences, Lines of text, Paragraphs	Gap-filling (cloze)
Drag and Drop	Total Deletion or Partial Deletion, Discrete item Deletion
Completion of tables	Text Expansion and Text reduction
Classification	
Jigsaw	
Hangman	
Crossword puzzles	
Selection/Ranking	
Question and Answer	
Multiple Choice Questions	
Varieties: T/F, only one right answer, several right answers	
Question Answer where correct answer matches pre-programmed right answer	
Designing a questionnaire as a student activity	
Matching	
Two halves of sentences	
Titles with texts	
Words and meaning	
Pelmanism (pairs)	
Grouping into sets	

The above taxonomy is limited as it does not take into account a more open ended reflective learning approach and may only work for stand-alone CALL courses or as a supplement to courses where teachers exploit

9

computer technology as a tool. These exercises are not sufficient for distance learning courses where contact with teacher or e-moderator is required. These exercises can be used to create simulations, mazes and games. Language learning simulations may combine a variety of these exercise types in an attempt to replicate some real-life activity. In mazes the learners are presented with adaptive sets of multiple choice questions at different stages of solving the puzzle. Games can be created by combining exercise types from this table with a stated goal and rules of play providing the learners the opportunity of winning or losing. The exercises listed in the table may exploit communicative language teaching's pedagogic principles but can be termed reductive in nature with 'yes/no' answers or minimal answers. These exercise types may give instant feedback or scores, yet they are limited because they are not open ended and do not encourage reflective learning (unless accompanied by class discussion or interaction with an e-moderator).

Most of these exercises are also used in paper-based materials developed for language learning. The potential value of CALL lies in its capacity to make them more dynamic and vibrant for students, compared to the impact they have within paper based materials. The level of interactive engagement changes with use of the technology-enhanced environment.

However sophisticated the technology, CALL exercises are essentially of three or four basic types: a) Text Manipulation, b) Gap filling or Text Reconstruction exercises, c) Question and Answer routines including multiple choice questions and d) Matching Sequences. The 'drag and drop' sequences that the more recent software is exploiting are still, in essence, Text Manipulation, Gap Filling and Matching exercises as answers proposed by learners are matched to a right answer. In fact all the exercises are in that sense, matching exercises because the learner's answers have to match the right answer pre programmed in the computer. Brett (1994:329) cites Scarborough (1988: 301) as noting then that 'a relatively narrow range of language-practice techniques [are] available ... just four types of program: gap filling, text manipulation, text reconstruction and simulation'.

The materials which will be evaluated in this study (EASE) employ most of the exercise types listed in Table 1.1 (with the exception of text reconstruction and games such as Hangman) and are subject to the same constraints. However, exercises in EASE also try to go beyond this taxonomy in their use of more open-ended questions with no absolute right or wrong answers, and ask learners to reflect on the experience of the activity.

1.2.3 Essential Academic Skills in English (EASE)

In this study, I shall engage students in using and evaluating the interactive multimedia software EASE (Essential Academic Skills in English) developed by the team at CELTE, University of Warwick. These materials constitute a series of interactive CD-ROMs designed for use by students whose first language is not English, who intend to study in Britain or another English speaking country, and who thus need to develop their skill in using English for academic purposes. The aim of the two CD-ROMs is to improve the students' academic listening and speaking skills, facilitating their participation in lectures, seminars and presentations.

Principles behind the development of the material accord with those acknowledged by Hémard (2003). The material exploits authentic classroom lectures and seminars (constructivist learning theory) and adopts a communicative approach to the development of listening comprehension and presentation skills. The two CD-ROMs of CALL MM materials used in this study are described in the following sections.

1.2.3.1 EASE Volume One: Listening to Lectures

This CD-ROM is the first of the series and aims at helping international students whose first language is not English develop academic listening skills. The CD-ROM has 85 short video clips derived from a data bank of lectures from twenty five different departments across a range of faculties in the sciences, social sciences and humanities. These clips (1-2 minutes long) are followed by listening and note taking exercises. There is an introductory 'tour', and six main units. These units focus on different aspects of academic listening and learning from lectures.

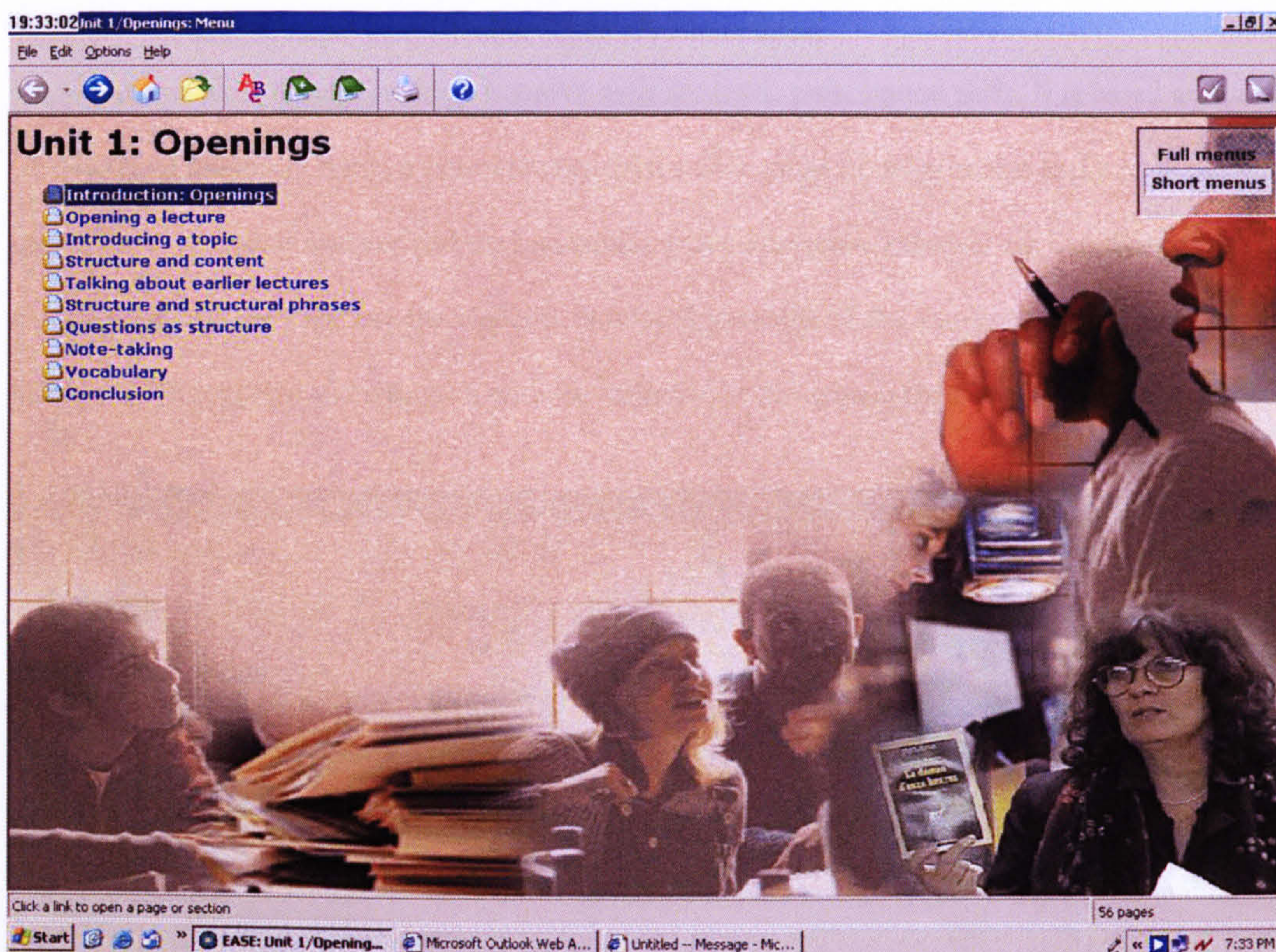


Figure: 1.1 A screenshot of EASE Volume: One *Listening to Lectures*

In the first section there is an introduction to the aims of the CD-ROM and an overview of the contents. The first unit deals with how lecturers begin their lectures by giving overviews or previews of lecture content. The second unit deals with the structure and organisation of lectures and demonstrates through video clips how lecturers use transitions and markers and give opening and closing summaries. The third and fourth units focus on language functions which include defining, classifying, exemplifying, comparing and contrasting. The fifth unit deals with significance and attitude where the students are taught to distinguish the more important points made by the lecturer from the less important ones. This unit has exercises on discerning the lecturer's slant on the topic, use of importance markers and differences between connotations and denotations. In the final unit of the CD-ROM students are familiarised with the concepts of thesis and anti-thesis, and their importance in academic discourse. There is a gradual progression in the difficulty level of the exercises from the first through to the sixth unit. The activities involve most of the exercises listed in section 1.2.2 with the exception of text deletion and games (Cf: Appendices 1.1 and 2.1).

1.2.3.2 EASE Volume Two: Seminar Skills (1) Presentations

'Presentations' aims to help students improve their academic presentation skills. It is based around digital video recordings of university seminar presentations from a wide range of departments and covers presentations from the sciences, the social sciences and the humanities. Many interactive exercises are based on these video clips. By working through the exercises the students develop knowledge of the features of an academic presentation, in terms of both language and structure, with the objective of developing their own presentation skills.

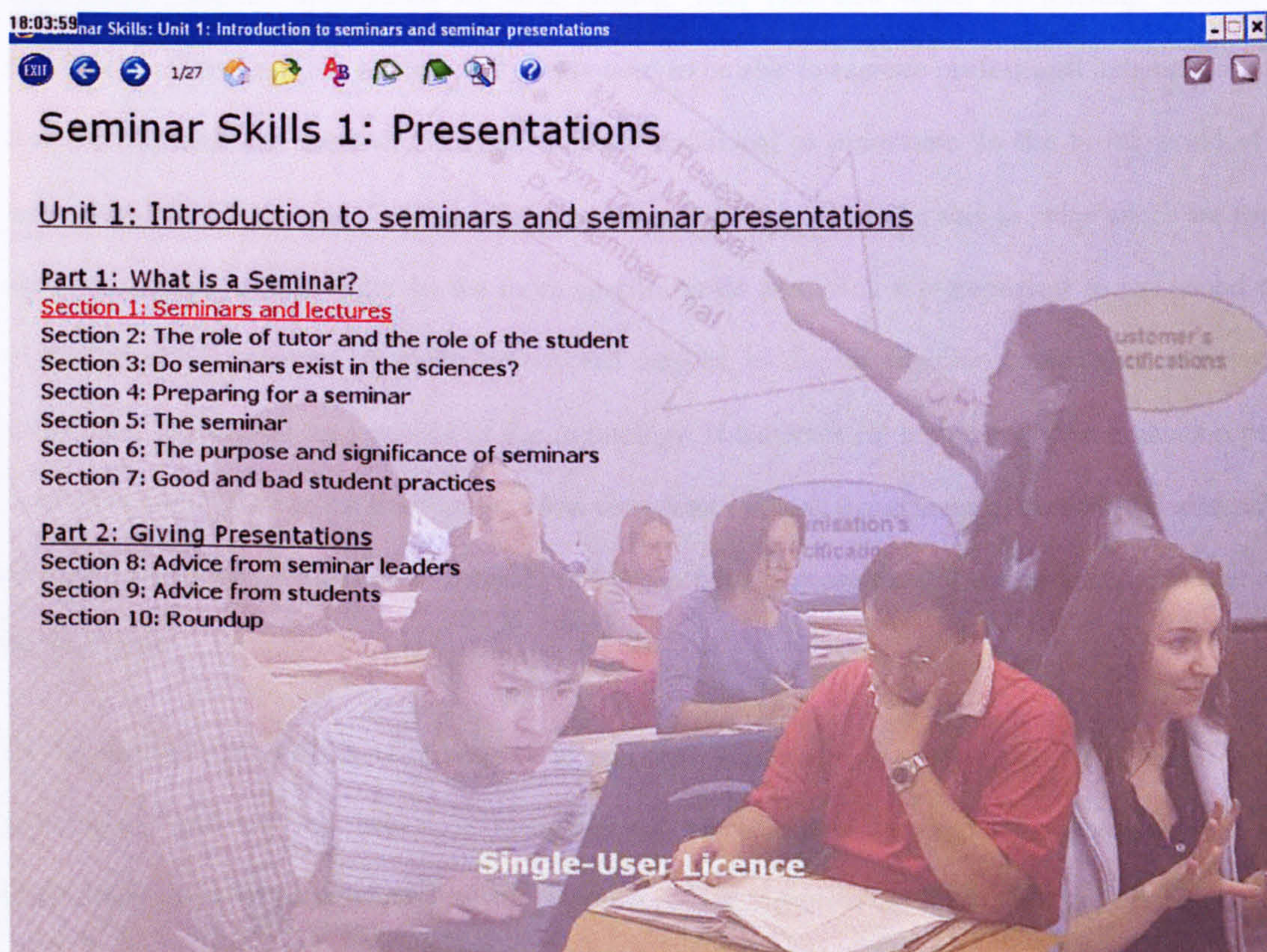


Figure 1.2: A Screenshot of EASE Vol: Two *Seminar Skills*

The main units discuss various aspects of making presentations. The first unit introduces the material and gives an overview of the CD-ROM. It gives a general introduction to presentations and seminars and highlights what lecturers expect from students. It emphasises the necessity of effective preparation prior to the presentation and gives practical advice on making presentations. The second unit deals with the importance of seminars and how to make good and effective presentations. The third unit gives practical tips about using audio visual aids, and how and when to use which aid. The fourth unit is about 'organisational signals' and gives practical advice on

how to overcome nervousness. The importance of referring to resources and proper referencing is given in Unit five, which also explains how to make statements seem more tentative and points out the difference between expressing views and giving opinions (Cf: Appendices 1.2 and 3.1).

1.3 The significance of this research

Learning technologies and the Internet have brought about a revolution in education, and multimedia/hypermedia computer applications are wide ranging. They raise new hopes for teaching, learning and the assessment of learning. But teachers will always need to be able to exercise professional judgment in relation to teaching materials and methods, whether they are traditional or innovative. In the broad world of learning technology, would-be users of multimedia materials or websites need to be able to judge which are trustworthy and offer content of real value. In the more specific world of CALL, it is important to use sound criteria in evaluating CALL software. A particular concern may be to discern whether materials are offering genuine innovation, drawing on the potential of the technology. If materials are innovative, then evaluation procedures need to be appropriate to the innovation. How then does a potential purchaser or user of MM materials reach a judgment about whether particular materials will deliver what they promise? What methods of evaluation should be employed?

In the field of CBI/HCI it is sometimes argued that weaknesses in evaluation of CBI occur because the procedures used are top heavy or management/ administrator focussed. Organisations and individual practitioners most often conduct evaluations which adhere to norm-based traditions i.e. drawing on checklists and questionnaires based on these checklists. In such evaluations, the evaluators check or rate the software according to the degree of agreement or disagreement with a given statement. They also draw on largely subjective written user reports. The weakness of these evaluations is that they are fragmentary; they deal with many details, but it is difficult to draw from them an overall impression of the materials. Some argue that their subjectivity, normative nature and lack of reliability make them ineffective. This kind of evaluation focuses on serving the organization which commissions the study, often at the expense of the learners. It would seem therefore that there is room for more learner centred evaluations, and a need also for objectivity to be achieved other than through normative questionnaires.

Although there is research on teachers' and students' attitudes towards CALL, little of it has taken account of students' specific insights and impressions when working on materials (Lasagabaster and Sierra 2003). For an evaluation scheme to be effective it needs to focus on the significant features of the learning materials, the learners and their experience of the learning task. Yildiz and Atkins (1993) suggest that evaluation should be student centred and content specific and should take into account students' familiarity with computer-based learning and their attitudes to it. Moreover, Hémard (2003:21) remarks that there is an emergence of new approaches "which are increasingly realising the importance and relevance of design considerations, alongside well established learning theories, to inform the design process, with a view to improving the conceptual underpinnings of design projects." Others also argue that evaluation of language teaching software is an important area of research as each evaluation has the potential to contribute to development in the field (Chappelle 1997:3; Felix 2003b:11; Yildiz and Atkins 1993).

New materials need to be evaluated and affirmed or rejected on the basis of empirical evidence. My interest lies in exploring these new approaches empirically and determining the appropriateness of various methods of evaluating MM CALL materials from a learner centred perspective.

It is potentially enriching for evaluation methods from the field of computer science to be available for use to evaluate the technical aspects of CALL software (Cf. Dix, Finlay, Abowd and Beale 2004; Karat 1997; Nielsen 1993; Nielsen and Mack 1994; Preece, Rogers, Sharp, Benyon, Holland and Carey 1994). However, CALL-specific methods of evaluation need to be alert to whether materials achieve the claims and intentions of the designers (closing the gap between intent and outcome) and whether materials are taking advantage of the full potential of the electronic medium. Do the learners use the materials in the ways envisaged by the materials developers? Is the practical use made of the materials consistent with the pedagogical principles which guided the developers, and which are apparent to an experienced language teacher on looking at the content and the proposed activities? This is of key importance.

A review of the literature identifies four approaches to MM materials evaluation, (Allum 2002; Kessler and Plakans 2001; Levy 1997; Scholfield and Ypsiladis 1994; Yildiz and Atkins 1993). The first approach is where evaluation is carried out by teacher researchers who generally base their inspection on experience and evaluate the software from the perspective of its general pedagogical approach and its availability, or the feasibility of

using it. Teachers may share the task if they particularly wish to avoid excessive subjectivity. A second approach involves experiment, where a comparison is made between groups using the materials in pursuit of a particular aim, and groups working without them in pursuit of the same aim. Such studies may be seen to determine the effectiveness of the materials and the way the teacher uses them. The third approach involves users' self reports or questionnaire responses, or uses key loggers to record errors and patterns of usage and choice. The fourth approach emphasizes evaluation from a technical viewpoint; and usability and functionality aspects of the program are considered (Lasagabaster and Sierra 2003: 294).

The purpose of this study is to explore and analyse the use of a certain number of different evaluation methods. Because of my commitment to the importance of user-centred evaluation, I decided to focus on methods from the third and fourth categories given their user-centred perspectives and the incorporation of usability- related evaluation perspectives. These methods will shift the focus from expert opinion and explore the experience and opinions of the students, the end users these materials are designed to help.

The main aim of the thesis is to identify a method or combination of methods capable of yielding reliable and comprehensive evaluation data on MM materials, drawing on learner experience, and signalling the potential impact and appeal of the software. In chapter two, the line of argument presented in this section will be more fully explored.

1.3.1 The study

The aim of this study therefore is to determine best methods and processes for evaluating CALL MM materials, bearing in mind current practice in the evaluation of language teaching materials generally, and the importance of a learner-centred approach. The study will explore different evaluation methods by getting groups of learners to use them in evaluating some materials. Identifying effective methods is important if potential purchasers of CALL MM materials are to be able to select suitable materials for their learners.

The central research question of this study is:

What is best practice in the evaluation of MM CALL materials and which methods or combination of methods can be employed for effective learner centred evaluation?

This key question gives rise to secondary research questions which are stated in the Chapter Three of this thesis (section 3.2.1).

During the course of the research a pilot study was conducted with the participation of 11 pre-sessional postgraduate students at the University of Warwick and a main study was conducted with the participation of 45 undergraduate students at the Lahore University of Management Sciences (LUMS). Students were involved in using two CD-ROMs in the EASE (Essential Academic Skills in English) series (see section 1.2.3). Students at Warwick, in the pilot study, used the materials for self-access study after regular class. An activity monitoring software and video camera were used to record their interaction and their views were elicited through the selected evaluation methods. The study at LUMS was carried out by integrating the MM materials into a regular course taught in a multimedia lab.

The students evaluated the MM materials at different stages of use, and data were also collected as the students used the materials. Altogether, evaluation methods used (and thus being trialled for their effectiveness) included: i) focus group interviews, ii) think aloud protocols married to user walk through and retrospective protocols, iii) pre- and -post use questionnaires and usability questionnaires and iv) observation and tracking of learner behaviour through auto key loggers and activity monitoring. Data gathered from these methods were analysed first as evaluation data, and then the method's performance was evaluated on a matrix.

1.4 The structure of this thesis

This introductory chapter has given an overview of the scope of the study, and the issues surrounding the development and evaluation of MM language learning software. The next chapter will review a range of literature concerning evaluation of computer-based materials and also paper-based materials, and will look at a number of previous evaluative studies of CALL materials, focusing on the methods they employed. There is detailed description of the various evaluation methods used, and an account of methods used more generally in software usability evaluation, which provided some inspiration for those employed in the study. The third chapter describes the research design, rationalises the methodology and presents the research questions. It gives an account of how the evaluations were conducted at the University of Warwick and the Lahore University of

Management Sciences. The chapter ends with the identification of a key set of evaluation criteria which are to be used as a means of measuring the effectiveness of the evaluation methods.

The next four chapters each deal with the process and outcomes of a different evaluation method. Chapter Four describes and rationalises the use of focus groups as a method of evaluating CALL materials, and analyses the data collected from student evaluation of the EASE materials by this method. There is discussion of the usefulness of this way of gaining student feedback on the materials, as used in this particular context, and a broader view of the effectiveness of the method is built up with reference to the criteria presented in the previous chapter, which are key to the thesis as a whole.

Chapter Five investigates think-aloud protocols (TAP) and retrospective protocols as effective evaluation tools. Cognitive user walkthrough is a related method encountered in usability evaluation, and this influenced the design of the TAP version selected for use in this study. The effectiveness of this method in externalising interior monologues and evaluative commentary is discussed from the perspective of the evaluator of EASE and with reference to the core criteria.

Chapter Six looks at pre and post questionnaires and usability questionnaires, and describes the use of questionnaires in the study. After consideration of the data they produced, they are also subject to analysis with reference to the criteria.

Chapter Seven discusses observation by key logging and activity monitoring. Again, the use of this method with students using the EASE materials is described, and again, the effectiveness of the method is discussed in relation to the established criteria. This method is something of a departure from direct learner-centred methods, since it focuses not on students' comments or elicited responses, but on observation of their actual use of the software.

Chapter Eight draws together the findings presented in the preceding four chapters, and presents an overall comparative analysis of the methods. Using the core criteria as a reference point final responses to the research questions are proposed, and the overall significance of the study is highlighted.

Having brought together this outline of the background, the aims and the methods of the study, and explained the structure of the thesis, we turn in the next chapter to a review of the literature within which the study is located, and to which it attempts to contribute.

CHAPTER TWO: A FRAMEWORK FOR THE STUDY

2.1 Introduction

There is a particular complexity in this study in that the focus of the research is on methods of evaluation, in order to evaluate the methods themselves in terms of their potential usefulness. I have tried to make this clear in the organisation of the thesis as a whole, and within chapters. This chapter and the next one, introducing a review of the relevant literature in the field and explaining the approach taken to the organisation of the study aim to make the complexity as clear as possible. A discussion of principles, practices and issues in evaluation in the broad domains of education, HCI, ELT and CALL together constitute the conceptual and theoretical framework and are reviewed in this chapter. My research procedure is described in Chapter Three with an account of how selected methods were operationalised in order to gain insight into the processes involved and to study their respective features, problems, and potential usefulness. This means that there is discussion of methods in both chapters. In the first, methods are part of the focus of the literature review; in the second, their operationalisation is discussed and described.

This chapter reviews the literature from four disparate yet connected domains in order to determine the shape and style of the practice of evaluation in these fields. The first section looks at the relationship between research and evaluation and traces the development of different approaches to evaluation. The second section looks at evaluation of materials in language teaching. The third section explores the practice of evaluation in HCI and the methods used for usability testing and evaluation. The fourth section reviews principles of evaluation of CALL MM materials, and recent CALL studies. After these have been drawn together, the final section reviews literature on the methods which are subsequently operationalised in the study.

2.2 Evaluation in education

2.2.1 *The relationship between research and evaluation*

Within the broad field of education, Cohen, Manion and Morrison (2003), Macdonald (1976) Mertens (2005), Norris (1990), Nunan (1992) Smith and Glass (1987) Stenhouse (1975) among others have explored the relationship between research and evaluation. Nunan (1992:193) affirms that "...any investigation which contains questions, data, and interpretations of the data qualifies as research." Therefore evaluation can be seen as the application of research methods to elucidate a problem or action. Looked at from this angle evaluation is not very different from research and instead becomes an extension of research, sharing its methods and methodology (Norris 1990:97). Mertens (2005:2) believes that the same methods of research are applicable to both lines of inquiry and adds: "In practice, a large gray area connects research and evaluation..." On the other hand some writers view research as a branch of evaluation – a branch whose mandate is to resolve the technological problems faced by the evaluator (Macdonald 1976:132). The distinctions between research and evaluation are considered by some to be blurred because of the areas of overlap (Smith and Glass 1987). Stenhouse is of the opinion that evaluation should lead development and be integrated with it "then the conceptual distinction between development and evaluation is destroyed and the two merge as research" (1975:122). These different opinions on the degree of overlap arise from different interpretations of the word "research."

According to Bennett (2003) the aim of research is the pursuit of new knowledge and therefore it can take different forms. One form is considered "pure" research which is open-ended and exploratory, seeking new patterns, explanations and theories. This is clearly distinct from evaluation and is more closely linked with the natural sciences than with educational research. A second type of research is "applied" research, which involves the testing of theories and hypotheses. In applied research the distinction between research and evaluation is not very clear-cut as it could be claimed that any innovation or new programme aimed at improving practice in education is a hypothesis about teaching, and evaluation tests that hypothesis (Bennett 2003 :15).

The ways in which research and evaluation may differ can be distinguished on the basis of certain characteristics. An evaluation study gathers information to judge a particular innovation with the purpose of informing decision-making (Smith and Glass 1987: 33-8). On the other hand, research aims to advance the

frontiers of knowledge to gain general understanding about the phenomenon being observed and researched. Moving on to differences in scope, Smith and Glass (1987) suggest that the research study is likely to have a narrow focus as opposed to a more comprehensive focus in an evaluation study. However, counter to this assertion it can also be argued that evaluation is about what works in a specific context and it has a narrower focus, whereas research has a wider scope as it purports to discover generalizable truths. According to Smith and Glass (1987) the parameters and the agenda of a study would be determined by the researcher in the case of a research study and possibly by a client commissioning the evaluation in the case of an evaluation study. Similarly the origins of an evaluation study could reside with a client who is commissioning the evaluation as opposed to the research study where the motivating factor would be curiosity and the researcher's need to know (ibid). The evaluator would be responsible and accountable to the client who commissioned the evaluation whereas the researcher would be accountable to the research community to which he or she belongs (Smith and Glass 1987:33-8). Evaluation takes place when a problem arises or a decision has to be reached whereas research is not time- or occasion - bound in the same sense, and can take place at any time.

Another difference between research and evaluation is related to the question of values. Research represents neutrality in values whereas evaluation represents the multiple values of the various stakeholders. The criteria for judging an evaluation study are its utility and credibility whilst for a research study the criteria are internal and external validity (Smith and Glass 1987:33-8).

However, whatever may be the similarities and the dissimilarities between evaluation and research, the literature affirms that those undertaking evaluation will have to draw on the strategies and techniques of research if they want to gather systematic evidence (Bennett 2003:15; Mertens 2005 :2).

2.2.2 Models and approaches in educational evaluation

“Model” and “approach” are terms that have often been used interchangeably in the literature on educational evaluation. Oakley (2000) builds on the work of Lawton (1980, 1983) Stake (1986) and Norris (1990) among others, and brings together a summary of the chief characteristics of the two main paradigms of educational research and evaluation. These two paradigms are the positivist scientific quantitative paradigm (henceforth described as “scientific”) and the naturalist, interpretive and qualitative paradigm (henceforth described as “naturalistic”). Oakley (2000:26-7) draws a distinction between the two paradigms on the basis of their aims,

purpose, approach, preferred techniques, research strategy, stance and method of inquiry. He explains that in the scientific tradition the aims of the evaluation are to test a hypothesis for the purposes of generalising the findings whereas in the naturalistic tradition hypotheses are generated through description of a phenomenon. The purpose in scientific inquiry is verification whereas the purpose in naturalistic inquiry is discovery. The scientific approach is top down and the naturalistic approach is bottom up, with the preferred technique of gathering data being quantitative in the scientific paradigm and qualitative in the naturalistic paradigm. The stance of the researcher in the scientific paradigm is reductionist, oriented to prediction and control, outcome orientated and exclusively rational. The stance of the researcher in the naturalistic paradigm is expansionist, exploratory, inductive, process oriented and intuitive. The methods in the scientific paradigm are ones that employ measurements and counting such as statistical records, surveys, experiments, case-control studies and content analysis. In the naturalistic paradigm the methods used include participant observation, in-depth interviewing, action research, case studies, life history methods and focus groups. The data revealed through scientific inquiry is hard, reliable and replicable whereas the data in naturalistic inquiry is rich, deep and valid. The relationship between theory and research in the scientific paradigm is that of confirmation while in the naturalistic paradigm it is emergent. The source of theory in the scientific paradigm is *a priori* and in the naturalistic paradigm it is grounded. The image of reality in the scientific paradigm comes across as singular, tangible, fragmentable, static and external whereas the image of reality in the naturalistic paradigm is multiple, holistic, dynamic and socially-constructed (Oakley 2000: 26-27). Evaluation in education has not strictly adhered to the tenets of these two paradigms but often has also mixed the two approaches.

Different models of evaluation bring very different perspectives to bear on the process of evaluation. The two most contrasting models, which clearly illustrate their adherence to the scientific and naturalistic paradigms, are the classical research model and the illuminative evaluation model (Bennet, 2003). Lawton (1983) Stake (1986) Oakley (2000) and Bennet (2003) concur substantially in their explanations of these two models and suggest that the classical research model is based on the scientific experimental model, which involves the testing of a hypothesis. It uses control groups and experimental groups which are given some form of educational treatment, or undergo an intervention, such as a new teaching technique which is applied to experimental groups. Both groups are retested and the performance of the group is compared to assess the affects of the

treatment. The classical research model places emphasis on the reliability and validity of the data that is collected. Lawton (1980, 1983) worked on developing it to accommodate different educational situations.

The second model of evaluation, the illuminative evaluation model was presented by Parlett and Hamilton (1976). They disagreed with the proponents of the classical approaches to evaluation on the difficulty of matching groups of students for experimental purposes in educational settings because of the need to control many important variables. They further suggested that for any specific context it is impossible to determine beforehand what all the relevant variables might be. (This becomes more acutely relevant when evaluating MM materials as the options and choices available to the students increase manifold in such materials). Their proposed approach drew upon the methods of social anthropology to study innovations in context and without the need for control groups. Unlike the classical research model, this approach takes into account the wider contexts of educational innovations and considers description and interpretation as opposed to measurement and prediction. This means that whilst data is collected to test a hypothesis in the classical approach, in illuminative evaluation a hypothesis is generated from within the data that has emerged (Parlett and Hamilton 1976: 88).

Parlett and Hamilton (1976) suggested a three phase model of educational evaluation, involving collecting open-ended data, exploring issues which have arisen, and identifying patterns. They suggest using four methods of collecting evidence: observation, interviews with participants, questionnaire and test data, and documentary or background information giving the context of the innovation. They see quantitative data as less significant and informative than qualitative and represent the outcome as a case study. The emphases on the context 'focus on the learning milieu' and the "notion of illumination" in the illuminative model produced quite a few followers amongst language program evaluators (Kiely and Rea-Dickens 2005:33).

Despite the flexibility of the illuminative evaluation approach to educational contexts and its popularity in the 1970s and 1980s, it has been criticised for being as limited as classical evaluation and for not having an adequately formulated "body of theory or methods" thus rendering it incapable of generating "a coherent, cumulative research tradition" (Atkinson and Delamont 1993: 218). According to Atkinson and Delamont the illuminative approach thus runs the risk of substituting one set of "atheoretical findings" derived from interviews and observations for the ones derived from test scores in the classical research approach (1993:218).

The critics of illuminative evaluation also raise a number of methodological objections regarding its use of case studies. One criticism is the influence of the researcher conducting the evaluation (Bennet 2003). The reliability, validity and objectivity of the data collected may be affected and there is a danger of the evaluators' views 'corrupting' the data with bias. To offset this concern evaluators can draw on multiple sources of data (triangulation) and clarify the stages in data analysis by providing summaries of all the steps taken in collecting and analysing data (data audit trails). A second concern is about the extent to which the findings of case studies can be generalised (Bennet 2003). However, proponents of the case study consider 'relatability' (Bassey 1981) and 'trustworthiness' (Lincoln and Guba 1985) as more important than reliability, validity and generalisability. The challenge lies in reporting the case study in such a way that its credibility is evident to people working in the field and they are able to relate to the issues being reported and are able to adapt and replicate the study to solve similar issues in their own context (Bennet 2003; Nunan 1992).

Other models of evaluation take less contrasting approaches to evaluation and some approaches combine the quantitative and qualitative aspects of these approaches. One of the early models called the 'countenance' model was proposed by Stake (1967) as a response to the narrowness and limitations of the classical model and included more descriptive aspects than the classical model. Another model born out of dissatisfaction with the classical model was "the goal-free evaluation model" proposed by Scriven (1973). This approach focuses on evaluating an instructional/reading package on the basis of the extent to which it meets needs. In this model a checklist is used to rate aspects of the instructional package in terms of its introduction, potential market and cost effectiveness. Scriven first used the terms 'summative' and 'formative' to distinguish between in-progress and end-of-programme evaluations. These two terms have since gained wide currency in evaluation literature including the domain of HCI, particularly in usability testing and evaluation (Scriven 1990; Dix *et al* 2004; Preece *et al* 1994).

Stenhouse (1975) adopted a different approach to educational evaluation, within which the teacher-researcher sees evaluation as an important element of curriculum development. Stenhouse gives teachers a crucial role in evaluation by saying that since evaluation rests on the work of teachers, teachers need to research evaluation and innovation themselves (1975:143). There is a considerable overlap between Stenhouse's approach and illuminative evaluation. This model was popular in the 1980s and 1990s with emerging work taking the form of

case studies or action research. Kiely and Rea-Dickins (2005:35), state that the “contributions of Parlett and Hamilton, and Stenhouse and his colleagues, in many ways rewired evaluation theory”. They go on to opine that “Stenhouse’s work has proved influential and enduring across British education generally and language education in particular’ (Kiely and Rea-Dickins 2005: 35).

Hargreaves (1996) observes that developments in educational evaluation show a trend towards more experimental approaches to evaluations because many countries face more centralised control of education and there has been an effort on the part of governments to raise standards. Due to this increased sense of accountability the educational research community has been encouraged to work at the approaches used in evidence- based medical research which tries to reach definite conclusions about what works (Hargreaves 1996: 5).

Another approach in educational evaluation is that of the “design experiment”. This approach draws on approaches used in technology and engineering which look at how a product, is designed to solve a particular problem, performs in a selected situation. Usability evaluations of software are inspired by this approach. According to Bennett (2003) this approach has direct relevance to educational contexts where the ‘product’ being tested could be a new educational package (curriculum, materials, self access computer based materials) developed to overcome certain selected shortcomings in a given context. It evaluates the effects of the new package in a limited number of settings. Design experiments have features of both the classical research and the illuminative evaluation approach to evaluation in that they seek to describe and explain what happens in selected situations and they also try to test a particular hypothesis (Bennett 2003: 39-40).

In the specific domain of ELT programme evaluation, both product and process oriented approaches have thrived, but in recent times the process oriented approach has seen greater currency (Lynch 1996; Rea-Dickins and Germaine 1998). This has resulted in a variety of evaluation methods being used in programme evaluation ranging from experimental design to social anthropology methods. Lynch (1996, 2003) draws a clear distinction between naturalistic/interpretivist and positivist/quantitative paradigms but has favoured an eclectic approach using both quantitative and qualitative approaches for his quasi-experimental study (Lynch 1990) in order to verify findings from multiple perspectives. Rea-Dickins and Germaine (1998) and Roberts (1998) have rejected experimental design altogether. The greater focus on action research in ELT has heralded ‘a stronger integration

of evaluation within practice (Rea-Dickins1994:84) and has resulted in greater involvement of the learners. This has paved the way for constructivist/ interpretivist paradigm to strengthen its foothold in evaluation practice discussed as *Fourth Generation Evaluation* (Guba and Lincoln 1994). The impact of this approach on evaluation methods is seen in the recent trend towards greater student involvement and collecting evidence of their learning processes by using self and peer assessment methods, portfolios and presentations (Ross 2003:7).

2.2.3 *Research strategies and techniques in educational evaluation*

According to Bennet (2003) the research strategies and techniques commonly used in educational evaluation draw on the sciences and social sciences. The five main strategies used in educational research are: action research, ethnography, survey, experiment and case study (ibid). Educational evaluation mostly employs experiments and case studies but where data is being gathered on a large scale in a number of locations, surveys are also used. When practitioner researchers are evaluating the effects of changes they have made in their practice evaluation becomes linked to action research (ibid: 55-56).

The five most commonly used research techniques are document study, focus groups, interviews, observation and questionnaires (Bennett 2003: 58). Other supplementary research techniques identified by evaluation literature in different domains including Social Science, Applied Linguistics and HCI can be audio and video recordings, field notes in ethnographic studies, photographs, key logger data, computer-screen shots, computer-activity monitoring and participant diaries. There is no consensus in the evaluation literature on how data should be gathered and what data should be gathered in an evaluation study. Different research strategies and techniques have their merits and demerits. Since the primary aim of any evaluation is to provide the best possible information there is merit in considering using different strategies and techniques and gathering both qualitative and quantitative data. The methods can be used in a number of possible combinations which also serve the purpose of triangulation. This discussion of research techniques is furthered in Chapter Three.

2.3 **Evaluation of language teaching materials**

The evaluation and development of materials are complementary activities (Hubbard 1996; Hémard 2003). Expertise gained from evaluating materials can be applied to the effective development of materials and the experience of development and currency with design principles can inform the process of evaluation of materials (Hubbard 1996). Materials evaluation has gained importance and complexity on account of the rapid

growth of new technologies which have impacted on the materials design and development industry (Levy 1997:41). Teachers also attach great value to the design, development and evaluation of teaching materials, as these are the tools of their trade. Free-market economies have also encouraged prolific publishing of teaching materials and because of this random abundance the need for sound and systematic evaluation procedures has become more pressing (Roberts 1996:376).

Language teaching materials can be published or produced in house by teachers. Published materials have already gone through some form of formative and summative evaluations before the end users (i.e. the teachers) undertake an evaluation to 'match' them to the particular needs of a context (Hutchinson and Waters 1987:97). The medium in which the materials are presented to the end user can be print, which includes textbooks, supplementary materials and worksheets, often supported by audio or video learning packages. Alternatively the materials can be on a computer disc or on the internet. Irrespective of what medium the materials are presented in, the reasons and purposes for evaluating materials are relatively few.

Different evaluators need to evaluate for different purposes. A group of teachers may evaluate materials for use in a particular context and may be fairly subjective. Individuals or group of individuals other than the teacher who will ultimately use the evaluation findings may also determine the selection of materials (McGrath 2002: 12). A ministry board may need to evaluate competing materials for inclusion on an officially approved list and would need to be more objective (so that the selection is acceptable all over a country). Both situations would need a set of criteria of evaluation based on a subjective and objective analysis of the context and the needs of the student body. The terms subjective and objective analysis with respect to evaluation of materials were first discussed by Hutchinson and Waters (1987:97-99).

2.3.1 Approaches to materials evaluation

The ability to evaluate teaching materials effectively is a very important professional activity for all EFL teachers (McDonough and Shaw 1993:63). There are very few teachers who do not use published course materials at some stage in their teaching career (Cunningsworth 1984). Materials are often seen as the core of a particular programme and are often the most visible representation of what happens in the classroom (McDonough and Shaw 1993 :64). As Sheldon (1988: 245) says "it is clear that course book assessment is fundamentally a

subjective rule of thumb activity and that no neat formula, grid or system will ever provide a definite yardstick". It is imperative however to evaluate systematically, based on a rational scheme. McGrath advocates taking a rigorous approach to material evaluation as so much is at stake "...so much can depend on making the right decision about materials that it pays (in terms of money and time) to be as rigorous as possible when evaluating" (2002:14). Breen and Candlin (1987), Cunningsworth (1984), McDonough and Shaw (1993) Roberts (1996) Sheldon (1988), and Williams (1983), all emphasize the need for a consistent and systematic approach to materials evaluation.

Evaluations can be carried out at three stages: pre-use, in-use and post use (McGrath 2002: 180) and they can use "the impressionistic method", "the checklist method" or the "in-depth method (ibid: 25-29) (These are discussed in greater detail later). They can be predictive or retrospective (Ellis 1997:36) and external or internal (McDonough and Shaw 1993). An external evaluation provides a brief overview of the materials from the outside, examining the cover of the book, the introduction, the publisher's blurb and the table of contents. Internal evaluation engages in a closer and more detailed examination (McDonough and Shaw 1993). These 'external' and 'internal' categories correspond to the "impressionistic method" and "in-depth" method mentioned by McGrath (2002:25-29). McDonough and Shaw (1993) argue that the evaluation process is never static and that the ultimate success or failure of material can only be determined after it has been evaluated in use.

Predictive (pre-use) evaluations which precede adoption of materials, are based on checklists, are study-specific and therefore "remain inexact and implicit" (Ellis 1997:36). There are fewer studies which identify ways and means of evaluating materials retrospectively (post- use). Ellis (1997:37) says "...there are a few published accounts of retrospective evaluations of course materials and very little information about how to conduct them". Retrospective evaluations would determine what aspects of the materials worked and what did not work and what had to be adapted to the context (Ellis 1997: 37). This kind of information makes in-use and post use evaluation methods particularly valuable. 'Trialling' or trying out with a target group is widely advocated in the evaluation literature especially when 'large scale or long term adoption' is being considered (McGrath 2002:13; Cunningsworth 1995).

We now turn to a discussion of evaluation methods and techniques.

2.3.2 *ELT materials evaluation methods*

This section uses McGrath's (2002:25) categories of "impressionistic method", the "checklist method" and "the in-depth method" (referred to earlier in 2.3.1) to discuss the merits and limitations of these approaches. These categories, however are not watertight and distinctions between methods are not neat (McGrath 2002:26). Moreover, this section also discusses Weir and Roberts' work (1994) who use a categorisation system which distinguishes between direct and indirect methods of data collection. The checklist method has been explained in more detail here than the other methods, but this imbalance can be justified on the basis of its prevalence and popularity in ELT materials evaluation.

2.3.2.1 *The impressionistic method*

Cunningsworth (1995), Hutchinson (1987), Johnson (1986), Lee (1975) and Stevick (1972) have discussed this method or variations of it.

In this method the idea is to obtain a general impression or an 'impressionistic overview' of the materials (Cunningsworth 1995:1). It looks at the treatment of particular language elements, the types of exercises used, and the authors' view of learning. Techniques of impressionistic evaluation cover a wide spectrum and look at different qualities, dimensions and components of textbooks. When these techniques become more specific and detailed they stop being "impressionistic" and move into the realm of "in-depth" evaluation. According to Ellis (1997) this evaluation is predictive in nature; McGrath (2002) tacitly concurs by suggesting that it be employed in the pre-use stage.

The strength of this method appears to be its ability to provide a quick overall impression and its practical value for busy teachers. A weakness of this method is that it gives the impression of superficiality even if the general impression of materials is gathered quite systematically as in Stevick's scheme (1972).

2.3.2.2 *The checklist method*

Tucker (1975), Williams (1983), Cunningsworth (1984), Sheldon (1988), Matthews (1991) and McDonough and Shaw (1993) have worked in the area of developing checklists which can act as criteria on which materials can be evaluated. Tucker (1975) takes a quantitative approach and Cunningsworth (1984) and Sheldon (1988) take a qualitative approach. Tucker (1975) used a scoring system but Matthews (1991) and McDonough and Shaw

(1993) do not advocate a scoring system. A more learner-centered approach is taken by Breen and Candlin (1987). Criteria or checklist based evaluation is seen by McGrath as part of “armchair evaluation” (2002:13).

This method has the advantages of being systematic, cost-effective and convenient because information is recorded in a convenient format for ease of comparison between competing materials. Because it is explicit it offers a common framework for decision making and the categories are clearly understood by all participants (McGrath 2002; Chambers 1997: 31). Skierso (1991: 440) discusses the systematicity of this method by suggesting that a checklist should consist of a comprehensive set of exhaustive criteria which are based on the basic linguistic, psychological and pedagogical principles which underpin modern methods of language learning. This is in keeping with Tucker’s suggestion that these criteria should be exhaustive enough to include the assessment of all aspects of materials and at the same time be discrete enough to focus attention on one characteristic at a time (1978 :219).

The limitations of this method are that each checklist is composed of elements which are relevant to the specific context for which it is being used and hence may not be suitable for another context (Ellis 1997; McGrath 2002). Checklists created by others have to be tailored to suit a particular context and this involves a lot more than just eliminating checklist items that are not applicable to the new situation (McGrath 2002:27). Another limitation of checklists is that they cannot be a static phenomenon; the categories of all checklists are a reflection of the time at which they were conceived and of the particular beliefs of their designers (Williams 1983).

McGrath (2002:33) identifies specific criteria for a checklist which can be used for ‘first-glance evaluation’. A shortened version of this list is given below in Table 2.1 as an exemplar of checklist items.

Table 2.1 McGrath's Checklist (2002)

Practical consideration		Context-relevance	
all components available?	Y/N	suitable for course:	
affordable? ...	Y/N	- length of course?	Y/N
Support for teaching and learning		-aims of course?	Y/N
-teacher's book	Y/N	-syllabus ? ...	Y/N

-tests?	Y/N	suitable for learners:	
- cassettes?	Y/N	-age?	Y/N
Likely to appeal to learners		-level?	Y/N
Layout	Y/N	-cultural background?	Y/N
Visuals	Y/N	suitable for teachers :	
Topics	Y/N	require resources available?	Y/N
		evidence of suitability (piloted in local context?)	Y/N

This kind of checklist has uses but does not relate to the pedagogical principles on which materials may be based. Other checklists do focus on pedagogical principles and a selection of items from a few such checklists in ELT (that McGrath cited (2002: 45) is given below:

Attention to grammatical accuracy

Balance of language skills (enough attention to reading and writing)

Enough roughly tuned input

Practice of individual skills integrated into practice of other skills

Plenty of authentic language

Encourages learners to develop their own learner strategies

(Drawn from: Harmer, 1991; Matthews 1991; Ur, 1996)

2.3.2.3 The in-Depth method

In-depth evaluation can help in determining whether materials live up to the claims of the publishers and authors (McGrath 2002). Cunningsworth (1995), Hutchinson (1987), Johnson (1986) amongst others have suggested different ways of conducting in-depth evaluations. The in-depth method involves a thorough examination of different aspects of the materials, (McGrath 2002: 27). The focus of analysis is on tasks and what is required of users, for example: what the learner has to do in the task, whether the focus will be on form, meaning or both, what cognitive operations will be required, what form of classroom organisation will be

involved, what medium will be involved, and who or what will be the source of language or information. The focus is also on the selection and sequencing of content and tasks and the distribution of information across teacher and students. This method has certain disadvantages despite being very thorough and considered. One disadvantage could be that the samples (exercises, tasks or units selected for analysis) may not be representative of the materials as a whole. Secondly it gives only partial insight because of its narrow focus. Thirdly it is time intensive and requires expertise which may not be available (McGrath 2002:27).

2.3.2.4 *Direct and indirect data collection methods*

Weir and Roberts (1994) distinguish between two kinds of data collection approaches for in-use and post-use evaluation: the direct approach which uses techniques like observations, document analysis, and test scores, and the indirect approach where data is derived from interviews and questionnaires. The two methods that these writers refer to as self-report methods of evaluation and data collection are interviews and questionnaires. Self-report data are indirect in nature as they either consist of a description of events from the perspective of an intermediary or represent the views of an individual, which cannot be directly accessed (Weir and Roberts 1994: 141). They go on to discuss that in educational settings direct evaluation data (documents observations, test scores) may be desirable but difficult to obtain (ibid). Furthermore, the perceptions of participants, however subjective are a crucial means to understand program implementation and effects, and such perceptions are only obtainable by self-report methods (ibid). Self-report data are unreliable if unsupported by other data because of post-event reconstruction by the informant and the tendency for interviewers or question wording to affect responses a fact fully realized in market research. Post-event reconstruction is the tendency to create an account of reality which is favourable to us as we have a vested interest in sounding good when we report what we have done. Low's (1991) analysis of conversational literature also suggests that there might be a reluctance to disagree in spoken interaction. Self report data is not unreliable when objective information about years of experience, age etc., is required, but can be unreliable when respondents are required to give opinions. There are methods available however to check the influence wording may have had on data, for instance when statistical tests are used for construct validity.

2.3.3 Concluding remarks

The above review of literature emphasises that a teacher led in-use or post-use evaluation of MM language teaching materials has to be rigorous and systematic. Moreover the purposes of the evaluation will be served best if care is taken to select methods that effectively elicit the kind of information that the evaluator is seeking.

Multimedia applications are different from paper-based materials primarily in their interactivity and the capacity to use animations, sounds and video along with text (Gyselinck *et al.* 2000). Because of the limits of the computer screen, the amount of information that it is possible to present at any given time is restricted. Various types of information have to be integrated and presented on successive screens in the case of complex phenomena (Gyselinck *et al.* 2000). To deal with these computer related features of MM materials it is in order to explore the practice of evaluation in the field of Human Computer Interaction (HCI). Perhaps evaluation of MM learning materials can harness some of these methods of usability evaluation to get the most meaningful data. The next section attempts a review of evaluation methods in HCI usability evaluation practice.

2.4 Evaluation methods in Human Computer Interaction

The International Organization for Standardisation (ISO) in Geneva, Switzerland has determined standards for measuring and ensuring quality of design relating to different aspects of computer software. These are extensively referred to in literature on usability inspection or testing (terms often used interchangeably with evaluation) in Human Computer Interaction (HCI) (cf. Dix *et al.* 2004; Karat 1997; Kirakowski 1998; Nielsen, 1993; Nielsen and Mack, 1994; Preece *et al.* 1994). Human Factors Standards deal with usability related aspects of human centred design. The “ISO 13407/DIS (1997) Human-centred design processes for interactive systems” is a standard that governs and guides human centred design, describing it as:

... a multi-disciplinary activity, which incorporates human factors and ergonomics knowledge and techniques with the objective of enhancing effectiveness and efficiency, improving human working conditions, and counteracting possible adverse effects of use on human health, safety and performance.

Literature on Human Computer Interaction (HCI) identifies certain design principles or guidelines on which hypermedia/multimedia language teaching applications should be based, and which help the user-interface designer throughout the design process (Dix *et al.* 2004:259). This knowledge essentially stems from three areas

of expertise: (i) the domain of cognitive psychology (Gardiner and Christie 1987) concerned with human-computer interactions such as *visual perception, attention, information processing, memory, learning, and mental models*, (ii) expert knowledge on data entry and display and user support, and lastly (iii) empirical data drawn heuristically from practical experience in user-interface design (Preece *et al.* 1994:492). To be effective, the guidelines have to be easily interpretable and usable by designers with no specific background in behavioural science (Dix *et al.* 2004).

Smith and Mosier (1986) presented guidelines for design of user-interface in six discrete functional areas: *data entry, data display, control of sequence, user guidance, data transmission and data protection*. The guidelines are derived from broad psychological principles through a process of simplification which filters, groups, interprets and translates them, through examples. They are then presented under fourteen “sensitive dimensions”: *design of procedures and tasks; analogy and metaphor; training and practice, task-user match; feedback; selecting terms, wording and objects; consistency; screen design; organization; multi modal and multimedia interaction; navigation; adaptation; error management; and locus of control* (Smith and Mosier 1986: 881).

After establishing how human factors feature in usability design we turn to evaluation of these features and methods which can be used to measure usability.

2.4.1 Usability evaluation

The definition of Usability used in ‘ISO 13407/DIS (1997) is as follows:

Usability: the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

Preece *et al.* (1994:722) defined usability as "a measure of the ease with which a system can be learned or used, its safety, effectiveness and efficiency, and attitude of its users towards it". According to Rosson and Carroll (2002:227) “a usability evaluation is any analysis or empirical study of the usability of a prototype or system”. Usability evaluation is considered an emerging and rapidly developing research area in the domain of HCI because it aims at improving the usability of multimedia software through formative and summative processes of quality control (Russell and Blake,1988; Shackel 1991:24). During the process of usability evaluation, experts consider the nature of users and the tasks they will perform under real world constraints in order to effectively

evaluate and improve multimedia software (Preece *et al.* 1994; Shackel 1991:24). ISO 9241 provides guidance on usability and emphasises “evaluating usability in terms of measures of user performance and satisfaction” (1995, 1999).

A review of the literature concerning Educational Evaluation and Usability Evaluation reveals that there is considerable degree of parity between the methods adopted by each as shown in Figures 2.1 and 2.2.

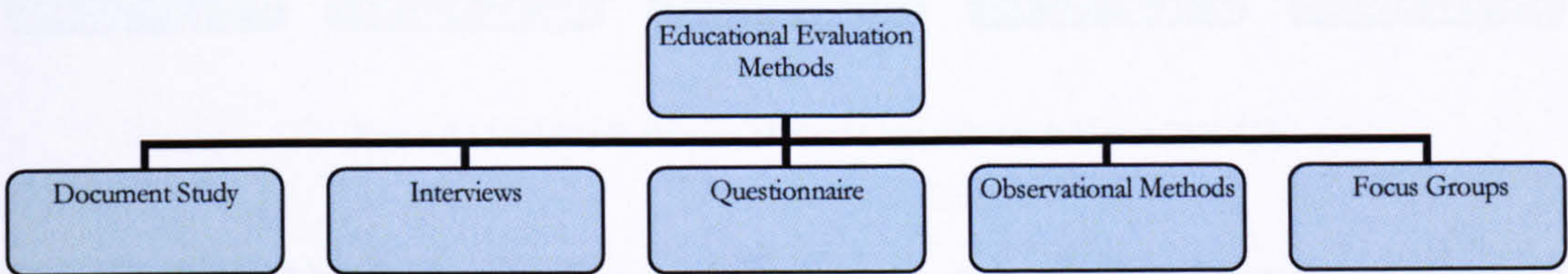


Figure 2.1 Evaluation methods in research (Bennet, 2003:58-59)

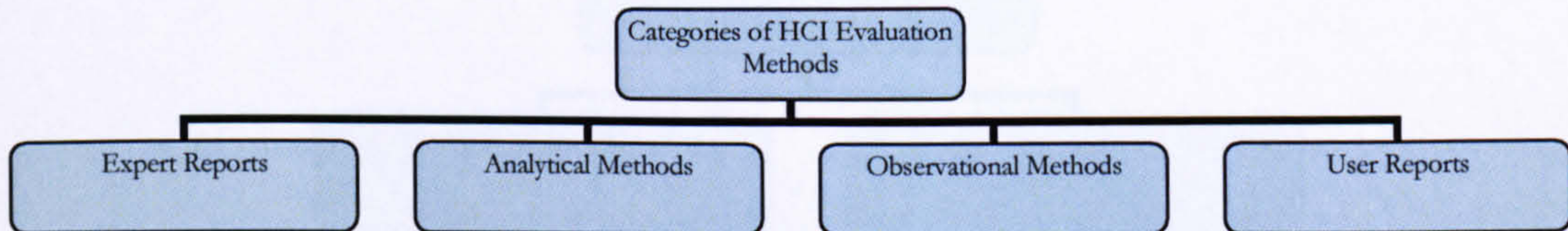


Figure 2.2 Usability evaluation methods in HCI (Whitefield *et al.* 1991)

Analytical methods in HCI which use heuristics or models to analyse the usability of software can be likened to checklists in evaluation of language teaching materials and *Questionnaires* in educational evaluation. *Document Study* where a close analysis of materials is undertaken corresponds to *Expert Reports* in HCI as the experts (software designers, engineers or developers) study the program being evaluated and their reports are analysed as documents. Similarly *User reports* aim to capture the same kind of information as *Interviews* and *Focus groups* are likely to in educational evaluation.

As can be seen in Figure 2.3, Ivory and Hearst categorise usability evaluation methods into five further domains of usability namely (i) testing, (ii) inspection, (iii) inquiry, (iv) analytical modelling and (v) simulation (2001: 473). Of these testing and inquiry involve users and experts alike whereas inspection, analytical modelling and

simulation require experts and software engineers and can be called expert based evaluation methods and are often done only by experts.

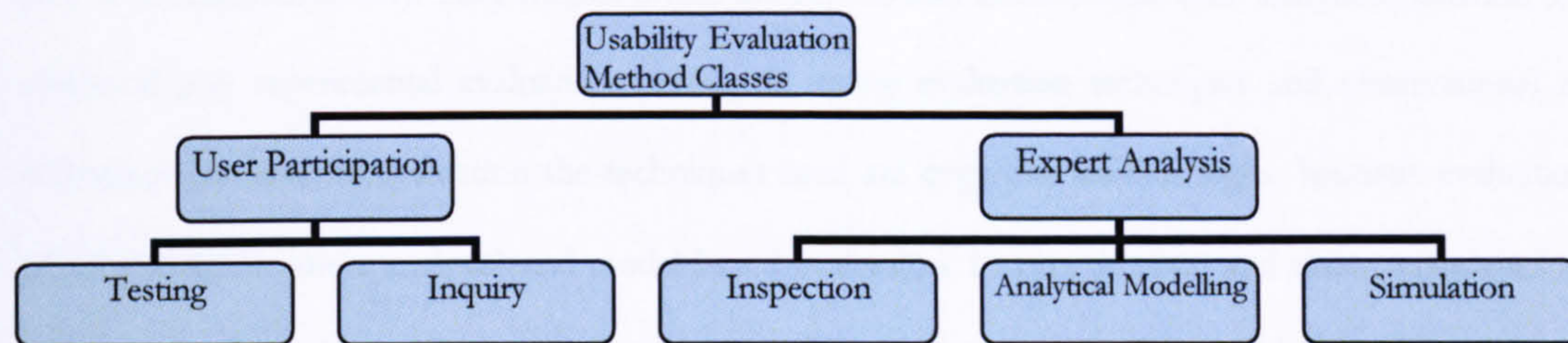


Figure 2.3 Usability Evaluation Method Classes (Ivory & Hearst 2001:473)

Dix *et al*, (2004: 319-362) also classify evaluation techniques under two broad domains of expert analysis and user participation. These correspond to the two distinctions of usability evaluations called ‘inspection methods’ which are done by experts and ‘testing methods’ which involve end-users.

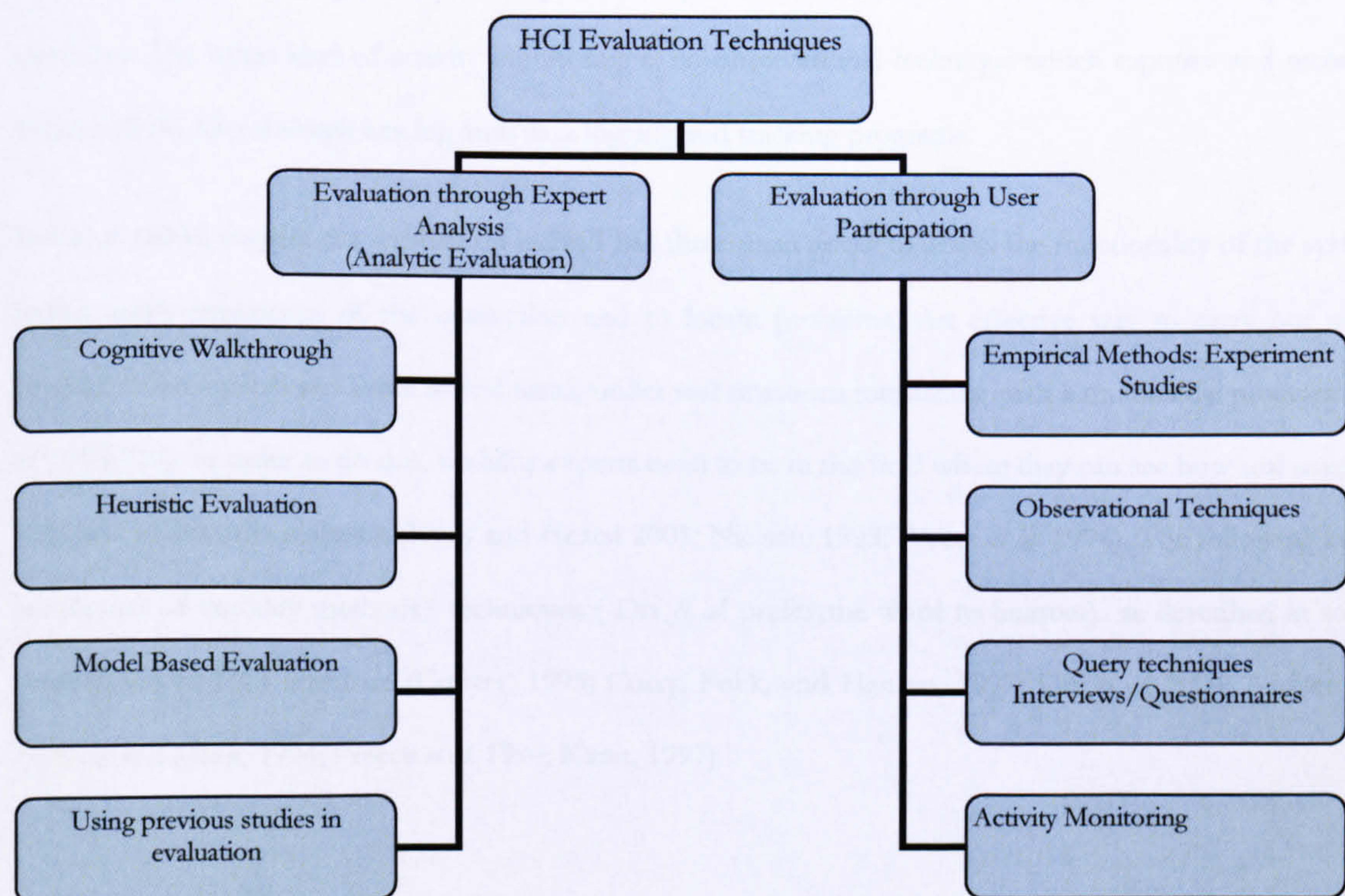


Figure 2.4 HCI Evaluation Techniques Using Experts and End-users

As shown in Figure 2.4 Dix *et al* (2004:360-1) divide methods into two broad categories of “evaluation through user participation” (on the right side of the diagram in 2.4) and “evaluation through expert analysis” (on the left side of the diagram in 2.4). They further divide the expert/user classification into: analytic evaluation techniques, empirical and experimental evaluation techniques, query evaluation techniques and observational evaluation techniques. In analytic evaluation the techniques used are cognitive walkthroughs, heuristic evaluation, review based evaluation (meta analysis) and model based evaluation. In experimental and query techniques they place experiments, interviews and observation. Lastly in observational evaluation techniques they place think aloud and post task walkthroughs (explained in Table 2.2.)

Concepts which may be new to educational evaluation or materials evaluation in Figure 2.4 are cognitive walkthrough, heuristic evaluation, model based evaluation and activity monitoring. These are explained in Table 2.2. Activity monitoring is of two types. The one mentioned here deals with using the computer screen as a camera /mirror or having cameras to capture the eye movement and other activity of the user (physiological measure). The other kind of activity monitoring is an observational technique which captures and records the actions of the user through key log auto data logging and tracking programs.

Dix *et al* (2004) suggest that evaluation in HCI has three main goals: to assess the functionality of the system, to assess user’s experience of the interaction and to locate problems. An effective way to carry out usability evaluation is to watch and listen to real users, under real situations interfacing with a multimedia program (Dix *et al* 2004:319). In order to do this, usability experts need to be in the field where they can see how real users work with real multimedia software (Ivory and Hearst 2001; Nielsen, 1993; Preece *et al* 1994). The following are brief summaries of usability methods/ techniques (Dix *et al* prefer the word techniques) as described in software evaluation and HCI literature (Conyer, 1995; Corry, Frick, and Hansen, 1997; Dix *et al*, 2004; Nielsen, 1993; Nielsen and Mack, 1994; Preece *et al* 1994; Karat, 1997).

Table 2.2 Evaluation Techniques in HCI Usability Evaluation

Method	Description
Observation	Observing users' behaviour throughout the evaluation . Could take electronic or paper pen forms. (Dix <i>et al</i> 2004; Nielsen 1993; Preece et al, 1994).
Questionnaire	Using questions concerning information and attitude about the usability of the software, the experience of the users. .(Dix <i>et al</i> 2004;Preece <i>et al</i> 1994; Karat 1988 : 897-898)
Interview/Verbal Report	Using user's verbal report while evaluation is underway or using interview or post event protocol after completing the evaluation (Dix <i>et al</i> 2004; Nielsen 1993; Preece et al, 1994).
Empirical Methods	Using a hypothesis to measure user behaviour while variables are manipulated by the evaluator (Dix <i>et al</i> 2004; Karat 1988)
User Groups	Using an organized (user forum/ focus group)or selected (beta sites) and learning from their experience (Nielsen 1993)
Concurrent Cooperative Think-Aloud	Using user's thoughts throughout the usability testing asking them to verbalise their strategies (Nielsen 1995 :375)
Cognitive Walkthroughs	Using a cognitive psychologist or a software expert to identify potential difficulties that a user may face through a step by step evaluation of a software design. (Dix <i>et al</i> 2004)
Pluralistic Walkthrough	Using teams consisting of different stakeholders in the design process (Preece et al, 1994).
Heuristic Methods	Using a team of experts to review a product or prototype to check compliance with usability principles. (Dix <i>et al</i> 2004 ; Nielsen 1993)
Review Methods	Using existing knowledge through a review of empirical and experimental studies in the research literature and industry standards(ISO) (Dix <i>et al</i> 2004)
Modelling Methods	Using of models like KLM (Keystroke Level Modelling) and GOMS (Goals, Operations, Methods and Selection) to predict and provide feedback on user interactions and difficulties. (Dix <i>et al</i> 2004)
Video Analysis Video camera/screenshots	Using videos to capture data about user interactions during usability testing. Chignell et al, 1995 :325) There are two kinds Video observation /computer based video or screen shots
Auto Data-Logging Program	Using auto-logging programs to track user actions throughout the usability testing. (Yoder, McCracken and Akscyn 1988: 871).
Software Support	Using software designed to support the evaluation expert during the usability testing process and to provide an evaluation summary. (Reiterer and Opperman, 1995 : 364-366)

Of these methods some can be used directly for evaluating MM language learning materials and some can be adapted. In order to understand what these usability evaluation methods can contribute to the field of CALL evaluation some of these methods are explained further. This selection is based on what the literature showed as most commonly used in HCI usability evaluation. These methods are not necessarily all user-centred methods but are a mix. However, this choice can be explained on the basis that there are possible parallels and parities between methods across disciplines and an understanding of these methods may help in identifying how evaluation studies have adapted and used these methods. Another rationale for this choice is the strategy

employed in this exploration which is to study the field first from a broad perspective in order to select a representative sample for more detailed analysis and then move towards a well informed final choice.

2.4.1.1 *Heuristic evaluation*

Heuristics are principles derived from different theories underpinning computer science and HCI theory which govern the design of software. The heuristic evaluation method was developed by Nielsen and Molich (1990). Nielsen (1994) recommends the use of his ten heuristics as they provide the most effective coverage of the most common usability problems. The method uses multiple evaluators (three to five experts or end users) who conduct independent inspections in which they compare interface elements with a predefined list of recognised usability principles (the heuristics) to find the usability problem (Nielsen, 1994:25; Preece *et al.* 1994: 676).

Heuristic evaluation has potential benefits in evaluation of educational multimedia. This method is similar to the checklist based evaluation methods used for paper-based materials. The advantages of this method are that it is easy to learn and use; is inexpensive to implement, is relatively speedy and can be used to identify problems early in the design process (Preece *et al.* 1994). This method can also be used for a summative evaluation taking place at the end of a program of software use in an educational setting. The limitations of this program are that participants need to be questioned and interviewed in a debriefing session to understand the software problems that were encountered, in order that they can be fixed (Conyer 1995). If this debriefing of three to five independent evaluators is not held there would be no way of knowing how these problems could be fixed.

2.4.1.2 *Cognitive walkthroughs*

In a cognitive walkthrough, a usability expert who knows the software questions the end user about his/her experience of working through the software According to Dix *et al.* (2004):

... the main focus of the cognitive walkthrough is to establish how easy a system is to learn. More specifically, the focus is on learning through exploration...So the checks that are made during the walkthrough ask questions that address this exploratory learning. (2004:321)

The main advantages of this method are that it is effective for predicting problems and hence it is useful for iterative evaluations (not the focus of this study). The cognitive walkthrough can be used for summative evaluations as well because it is effective for capturing cognitive processes (Wharton *et al.* 1994).

The main limitations of this method are that it needs a trained and skilled evaluator and is time consuming and tedious (Conyer 1995). Moreover to avoid confusion a cognitive walkthrough has to be focussed on one attribute of usability otherwise debugging may not be very effective (Dix *et al.* 2004:321).

2.4.1.3 *Pluralistic walkthroughs*

As the name suggests this method tries to access information about a product from multiple perspectives. The main aim of the walkthrough is to evaluate a product from the perspective of the end-user (Preece *et al.* 1994:679). The software designers and developers including the HCI engineers work together with end users on possible task scenarios (Bias 1994: 65). The main advantages of this method are that it is easy to learn and use and can be set up to meet the criteria of all parties involved in the evaluation (Preece *et al.* 1994: 679). It is considered most beneficial for iterative testing and evaluation (Conyer 1995). The main limitation of this method could be that it does not evaluate the program in the context it is likely to be used (*ibid.*). Conyer (1995) further adds that because this walkthrough is narrowly focused on one usability specific problem at a time it does not identify general problems.

2.4.1.4 *Formal usability inspection*

Expert reviewers conduct formal usability inspections by using models within clearly defined goal-oriented scenarios. Conyer (1995) uses the term 'formal usability inspections' while Dix *et al.* (2004) refer to something similar as 'model based evaluation' (as shown in Figure 2.4). The main advantages of these inspections are that they can be used to represent different knowledge domains. They can also be used to derive a list of problems and solutions for usability, and to evaluate both cognitive processing and behavioural tasks within the software (Conyer 1995). The limitations of this method are that the end-users are generally not involved and it is difficult to find a proper testing context for the tasks performed (*ibid.*).

2.4.1.5 *Empirical methods*

These methods involve hypothesis testing and are similar to the experimental studies mentioned in section 2.5.2 which are used in the domain of educational/usability evaluation and natural sciences. The advantages of using empirical methods for usability evaluation are that they can prove effective for finding cause and effect (Conyer 1995; Ivory and Hearst 2001; Nielsen 1993; Preece *et al.* 1994), and they can be instrumental in addressing a specific question or problem. The limitations of these methods are that they are time-consuming and expensive

to conduct. Moreover a trained skilled practitioner would be required to get the best results (Ivory and Hearst, 2001; Nielsen, 1993; Preece *et al* 1994).

2.4.1.6 *Formal design analysis:*

This method tests understanding of the task requirements to be performed. The steps required to solve a problem or perform a task are broken down into constituent steps and rated algorithmically and numerically by comparing to other design alternatives (Preece *et al* 1994). It can be useful for comparing the different design elements of usability. Another advantage is that it is effective for identifying problems at an early stage (Conyer 1995). The main limitations of this method are that it is difficult to learn and use and is best left to be used by experts (Ivory and Hearst 2001; Nielsen 1993; Preece *et al* 1994).

2.4.2 *Concluding remarks*

This section has reviewed evaluation methods enjoying currency in the field of HCI which may be relatively new to evaluators of CALL materials. In the next section, I turn to review a number of evaluations of CALL materials which are significant in the field, and which have drawn on a range of the methods dealt with in this chapter so far.

2.5 Evaluation in CALL

The use of multimedia in education is seen as an extension of computer based learning (Bates 1994). This is not surprising as it takes learning through computers a step forward by employing graphics, audio and video enhancements. Multimedia instructional materials are the latest state-of the art form of computer based teaching and learning materials. However, although evaluation is an essential step before materials can be adopted, teachers have shown a tendency to shy away from this task and have simply accepted whatever is available in the teaching packages (Bates 1997). Perhaps this is because of a lack of awareness of the procedures for evaluating multimedia materials or because the use of technology has not reached the 'normalisation' stage where all stakeholders become equally comfortable and conversant with it (Bax 2003:23).

'Multimedia' has been defined in the literature in many ways which can give rise to a conflicting understanding of the term, but a usefully straightforward definition is:

‘Multimedia is any combination of text, graphic art, sound, animation and video that is delivered by computer. When you allow a user- the viewer of the project- to control that and when these elements are delivered, it is interactive multimedia. When you provide a structure of linked elements through which the user can navigate, interactive multimedia becomes hyper media’ (Vaughan 1993: 3)

This thesis explores the question of what is the optimum method of evaluating MM electronic materials. Research has established that it is more difficult to teach a language fully online than other subjects in the humanities, it requires more resources and effort on the part of designers and teachers (Felix 2003a). Some argue however that the difference in the mode of delivery of interactive materials does not suggest a paradigmatic shift in learning and teaching (Salaberry 1996) and therefore need no ‘special’ approach to evaluation.

I am arguing however that in computer science and related disciplines and in the area of educational technology, many evaluation methods exist which also have potential for use in evaluating the technical aspects of CALL software and these and others merit research. This study aims to identify methods or a combination of methods which could lead to a reliable and comprehensive evaluation of not just the technical features of the software but also its pedagogical features.

In this section, I present an account of important work by Hubbard on how evaluation of CALL materials relates to design principles and also some published CALL evaluations.

2.5.1 The nature of Hubbard’s work

The evaluation scheme particular to CALL discussed here is based on a framework described by Hubbard (1988, 1992, and 1996). It is not an experimental study or an evaluation of CALL materials (like the ones discussed next in section 2.5.2) but a framework for the design and development of CALL materials. Hubbard (1996) suggests three stages for the development of courseware namely: Development, Evaluation and Implementation. Hubbard proposes that “Evaluation can in some ways be considered the inverse of development” (1996:27). From this we can conclude that he thinks that design principles can also act as evaluation guidelines.

There are three major components in Hubbard's framework: *Approach, Design, and Procedure*, (1996: 20-25) as shown in Figure 2.5. The 'language teaching approach' of a particular program or set of materials will depend on two elements 'linguistic assumptions' and 'learning assumptions' (Hubbard 1996: 20). How the program or courseware developer understands language and its structural, social and cultural aspects will act as guiding principles for him and can be referred to as linguistic assumptions (ibid). Linguistic assumptions may be based on theory but learning assumptions are based on how the developer understands the process of second language acquisition and the role of the environment on learning a language. The approach is also influenced by the capability of the computer itself (Hubbard 1996:21). The newer the machine the more varied its capabilities for language teaching purposes. Consideration of the computer *delivery system* can be combined with the developer's language teaching to formulate a set of design criteria. Hubbard (1996:21) provides a sample of humanistic communicative approach criteria. According to these criteria the courseware should provide (a) meaningful communicative interaction; b) comprehensible input at a level beyond that already acquired by the student; (c) opportunities for the student to develop a positive self image; (d) a challenge without generating anxiety or frustration; and (e) a catalyst to promote learner-to-learner interaction in the target language (Hubbard 1996: 21-22).

Hubbard's framework is effectively operationalised by D. Healey for evaluation of CALL software with the formulation of a set of questions on each aspect of Hubbard's model. She explained to me how this work was part of a teaching handout she used which Stevens published on the web (*Software Selection, Evaluation and Use*. Stevens 1996) [Deborah Healey, personal communication, 20 February 2006]. Figure 2.5 shows the aspects of Hubbard's model.

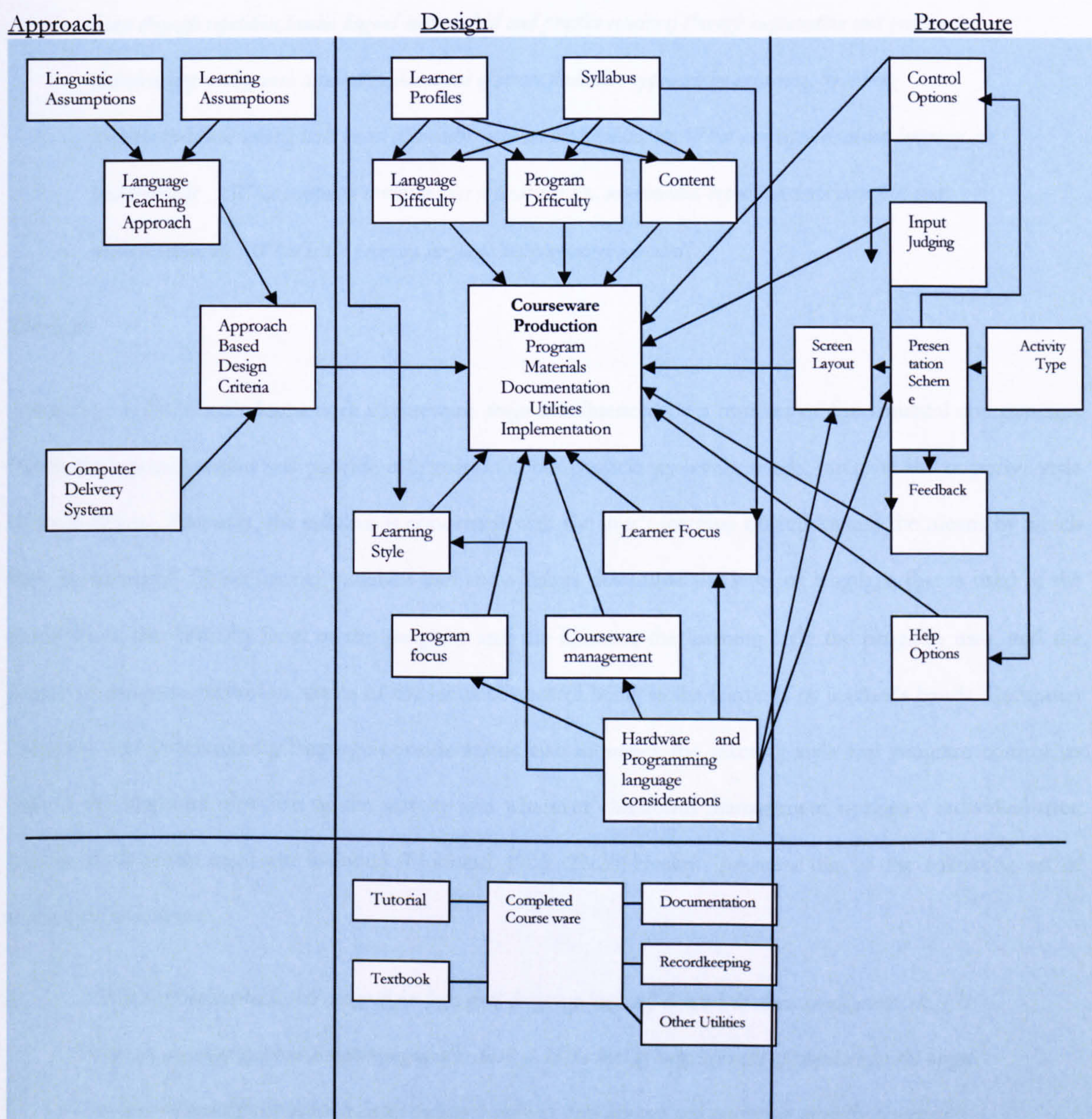


Figure 2.5 Hubbard's Development Module (1996:19)

Approach

According to Hubbard's framework the *approach* concerns the designer and the teacher/language practitioner more than the user-interface design expert or the end user, the learner, but analysing and determining the developmental process approach would be in the interest of the evaluator as well. So in the checklist there could be questions relating to the design of the software. Healey explains Hubbard's framework for evaluation of CALL software with the formulation of the following questions.

What language learning assumptions are embodied in the software? These might include assumptions that people learn through repetition, (audio lingual method-drill and practice routines) through explanation and practice (deductive approach), with detailed explanations of errors, (inductive approach by exploring, by solving problems (problem solving task based approach), with multiple media, etc. What assumptions about language are incorporated? ... What computer strengths does it focus on; e.g., multimedia, repetition, error analysis, text manipulation, etc.? What is the primary language teaching approach used?

Design

According to Hubbard's framework courseware *design* is influenced by a number of fundamental components. Firstly the learner profiles will provide information about proficiency levels, needs, interests and cognitive style of the learners. Secondly, the syllabus is concerned with the user's learning objectives and the means by which they are achieved. These learner variables and the syllabus determine the level of language that is used in the courseware, the difficulty level of the program and the content, the learning style the program uses, and the degree of program control in terms of the locus of control being in the learner's or teacher's hands. Computer hardware and programming language considerations also influence the learning style and program control, as well as the linguistic objective of the activity and whatever classroom management options (individual user, paired, small group etc.,) are included (Hubbard 1996: 22-23). Healey presents this as the following set of evaluation questions.

What type of learner is this designed for (age, goal, language level, self-directed, in class, small group, etc.)? Is there an apparent syllabus-a sequencing of activities? ... Is the level of language used appropriate for the target audience? How difficult would it be for the target audience-both learners and teachers-to learn the program? ... What learning styles are favoured by the program (visual-textual, visual-graphic, auditory, kinaesthetic)? How much control does the learner have? For example, who sets the amount of time, number of items, type of error correction, and sequencing of items? Is record keeping part of the program? If so, what is tracked and for how long? Can it be customized?

Procedure

The *Procedure* section incorporates the elements to be considered in the layout of the courseware. Decisions have to be made regarding what kind of activity types to include (e.g., game, text reconstruction, simulation or quiz), how material is to be presented, input judging (none, with feedback, etc.) the control options, (e.g. at what points and to what degree the learner controls what the program does) the help options and the screen layout, including audio and video components (Hubbard 1996:23-25).

What types of activities are included (tutorial, quiz, game, text manipulation, simulation, problem solving, and exploration)? How is the information presented (e.g., in text, graphics, sound, video)? What type of input judging is there? Does it include ways to screen for spelling errors? Does it allow alternative correct answers? When and how can the learner get help for the content and the operation of the program? Is the screen layout appropriate? motivating? understandable? Can you control the sound level? Does it work fast enough over the network? [D.Healey]

Such questions can be based on criteria predetermined by the evaluators. The criticism levelled against the checklist approach for paper-based materials also pertains here, however. Such checklists are time, product and context specific, particularly in the case of electronic materials because technology is changing so rapidly that a checklist designed in 1992 may not be entirely applicable in 2002 or 2006. Moreover adaptation must involve more than changing a few of the criteria in the checklist. Hubbard's methodological framework by his own admission does not take into account CALL methods but focuses on specific software. Since "...CALL is not a mature field" his conception of methodology does not relate to "what has been produced but ... what could be produced" (Hubbard 1996:16). So this methodological framework is for hypothetical software, or CALL materials that could be produced.

Hubbard does not tell his readers how to evaluate ("... the framework does not specify any particular evaluation process...") but just suggests 'matching' development criteria to evaluation criteria which he describes as 'learner fit (design) and teacher fit (approach)' (1996:26-27). So this framework, although helpful, does not really tell us which methods to use for a comprehensive learner-centred post-use evaluation. Nonetheless Hubbard's

model is valuable because it covers the whole range of processes for design development and evaluation of CALL software and to my knowledge only one of its kind in CALL literature.

When a program is evaluated by experts in the field before it is given to the learners it usually employs the evaluation checklist approach and can be classified as a predictive evaluation (Squires and McDougall 1996). According to Squires and McDougall (1996) checklist based reviews of educational software are riddled with problems about how much emphasis or weight is to be given to each question, whether evaluation of software of the same category may focus on similarities more than differences, whether there is greater emphasis on technical rather than educational issues and whether consideration is given to innovative software evaluation. Moreover these evaluations do not take into account differences in teaching styles and do not address the issues related to follow up tasks and extension activities that the instructor may use. Different subject areas require unique set of criteria. In sum, checklist approaches are problematic and subjective with a tendency to become judgemental, because experts may expect the software to measure up to their preconceived expectations, and learners may not be recognised as part of this 'equation'.

2.5.2 CALL evaluation studies and their methods

The studies critiqued in this section describe the evaluation of CALL materials, some of which are ELT materials, others for teaching other foreign languages. The purpose of the review is to establish the current situation in terms of existing published evaluations of CALL materials. This is of key importance in establishing the framework for the study I am undertaking. Most of these studies are in-use /post use evaluations undertaken to improve the design of the product. Although there is great emphasis laid on conducting systematic and rigorous in-use and post use evaluations of ELT materials (McGrath 2002) in reality teachers mostly carry out pre-use evaluation of materials, and that too at the level of what McGrath calls 'first glance evaluation'. MM materials cost a lot more than textbooks, however, and this makes it even more important for teachers to conduct in-use and/or post-use evaluations in order to make decisions that will stand the test of learner use.

2.5.2.1 *Legenhausen and Wolff*

Legenhausen and Wolff (1990) report on the evaluation of two CALL programs. They evaluated a language teaching simulation GRANVILLE (French) and a text reconstruction exercise using STORYBOARD. Their aim was to study the effectiveness of these materials and determine what strategies learners used in working through these CALL materials. They used two methods: (i) Verbal reporting for the text reconstruction exercise and, (ii) Video observation for the simulation. The *computer function* evaluated in the text reconstruction exercise was the computer's ability to function as a language tutor. The evaluative principle was 'cognitive psychological' because the researchers wanted to identify what mental activities learners engaged in and whether these mental activities were language learning processes. Legenhausen and Wolff used verbal reporting/think-aloud protocols which they considered as belonging to the domain of cognitive psychology and hence capable of eliciting learner strategies. The computer function evaluated in the second program, the simulation, was the computer's ability to simulate reality, and the evaluative principle was 'discourse analytical'. The questions being explored were: (i) whether intensive communication occurs between participants, (ii) what is the quality of the interactions in terms of 'discourse authenticity' and 'acquisitional value', and (iii) to what extent the computer facilitated the participants' identification with their simulation roles'

In the STORYBOARD exercise, experimental data was collected from two groups of respondents. One group (A) had never seen the text to be reconstructed before and the other group (B) had seen the text for 30 seconds. Both groups had 20 minutes to reconstruct the 103 word text. While the students worked on the text reconstruction exercises they also reported (verbal reporting) on the strategies they were using to guess the answers. Legenhausen and Wolff compared the two groups' reconstruction scores, arrived at by counting the number of correct and incorrect identifications of content and functional words, and identified different kinds of strategies, which they classified as: (i) text-independent (TI) strategies(9 in number), (ii) text-dependent (TD) strategies(22 in number), and (iii) memory strategies. The two groups employed a total of 385 TI strategies and 759 TD strategies. Only Group B informants remembered words, phrases and whole sentences of the text and no calculations or further detail is given about memory strategies.

Legenhausen and Wolff state their reasons for using verbal reporting as an elicitation technique. However, they do not report on the procedure of how it was undertaken but we can assume tape recorders were used and

accounts were later transcribed. It is not clear whether their classification scheme was tested for inter-rater reliability. They do not report on how easy or difficult their informants found it to verbalise their mental processes while doing the text reconstruction exercises, nor do they tell whether the informants had to be prompted to record their mental processes. The reader also cannot discern whether the use of the method actually impacted on the reconstruction scores, thus raising concerns about the validity of the findings. Considering the sheer number of strategies that they identified through verbal reporting does raise the question of how intrusive the method was in eliciting the information in the experiment.

This quantitative method of analysis was probably quite appropriate for a short text (103 words) used for reconstruction. It is worth bearing in mind, however, that the hypermedia nature and multimedia functionalities of modern CALL MM materials may not be so suited to Legenhausen and Wolff's experimental design. The full potential of verbal reporting as a method appears not to have been exploited in this study, and perhaps the method would have revealed richer findings if the data had been treated more qualitatively. For instance, there is no way of knowing how evaluative the informants were of their own performance and whether they knew cognitively what strategy they were employing. There was also no way of establishing whether language learning had taken place as pre-tests and post-tests were not employed.

Legenhausen and Wolff (1990) used video observation as a data gathering technique for their second experiment on GRANVILLE. This requires learners to imagine spending a five-day holiday on the French Atlantic coast, and plan their activities and budget on the basis of information available on the computer screen. Three students were given 45 minutes introductory lesson to familiarise them with the simulation, and then their interaction in the 60-minute session was recorded using two video cameras. The video data was transcribed and codified using a codification scheme which incorporated 3 different parameters: (a) discourse level, (b) interactional structure and (c) speech acts.

By means of these procedures, they concluded that GRANVILLE failed to fulfil the objectives of simulations as the students did not identify with the roles in the simulation. This, they thought, was, because the computer screen stayed central to the process, preventing interaction and eye contact between participants. Moreover the respondents also constantly switched discourse levels and the program relied heavily on multiple choice procedures. An overall impression of the effect of the program could have been gathered fairly quickly just by

observing the video footage, however, and might have enabled Legenhausen and Wolff to arrive at the same conclusions without the detailed coding and analysis of video transcripts that was undertaken.

Of the two methods used for evaluation, video observation appears to have been less intrusive, suggesting that the results of the second experiment may have been more valid than those of the first experiment.

2.5.2.2 *Desmarais, Laurier and Renié*

Desmarais, Laurier and Renié (1998) studied navigation patterns of learners working in MM learning environments. Their study used think-aloud protocols (Ericsson and Simon 1987) to identify navigation patterns. The research was conducted on the program *Vi-Conte*, using twenty six learners working in three one hour lessons of French as a second language. *Vi-Conte* is based on a Canadian short silent animated movie and aims to develop listening skills, expand vocabulary and familiarise learners with the French-Canadian culture in French as a second language context. It has three menus: the first menu, 'Observation', allows the learner to watch the original movie and answer questions or learn vocabulary related to the situation illustrated. The second menu 'Narration' presents the movie dubbed with a narration in French and includes comprehension questions and traditional grammar or vocabulary exercises. The third menu 'Réflexion' consists of slides and texts depicting different aspects of Quebec cultural life.

Desmarais *et al* (1998) differentiated between linear and chaotic patterns of navigation. When a student follows the program's order and completes the activities in sequence then the pattern is considered linear. When there is no obvious order and the activities are not completed then the pattern is considered chaotic. They observed that a chaotic pattern resulted when the student used ineffective strategies, suffered technical problems or found the content difficult. They also observed that the first lesson was more chaotic as the student spent time exploring the program.

Criteria were used to analyse patterns and sub groups were compared according to the variables of age, gender, personality type and level of French. One interesting finding in opposition to earlier work on the topic was that males used more linear patterns than females. Males also completed the activities while females would often stop and repeat. Intermediate level students presented more linear patterns than elementary ones leading the researchers to question the relevance of *Vi-Conte* for beginners.

Desmarais *et al's* (1998) work is based on earlier studies by Qui (1994), Schwier and Misanchuk (1993), Horney (1993), Orey and Nelson (1994) and Recker (1994) on the linearity/non linearity of navigation patterns. In contrast to Desmarais *et al's* findings Recker's (1994) work on hypertext -based materials found that novices used more linear navigation with little or no backtracking and more advanced and perhaps experienced learners preferred a more non-linear top down approach because they were more goal oriented.

In the same paper Desmarais *et al'* (1998) also report on the learner navigation patterns from another program *Camille*. They state that the specific nature of the observations and findings from one program *Vi-Conte* does not allow them to be considered as general tendencies that could guide the development of other environments. They think that in order to generalize results, additional studies with other multimedia language learning programs are necessary. They applied a similar methodology to the program *Camille*. The module they used illustrated different communicative situations followed by various exercises to learn how to sell a product in French. In this research, instead of using the transcriptions as raw data, they used an 'on-line tracking' program (auto data logging program in Table 2.2) which recorded each learner's actions along with the time used. This data appears to have been easier to prepare for analysis and, because the entry is automated, good reliability is ensured. Using the database program that interprets action codes, they recreated synopses of the sessions. These synopses were then used to draw navigation patterns. The validity of this method will stay high as long as it is just navigation that is studied if the researchers want to use this data to infer meaning from these patterns, triangulation with more qualitative methods will become necessary. However despite these concerns Desmarais *et al's* (1998) study clearly indicates one future direction of CALL evaluation as likely to involve researching navigation patterns through automated observation to understand the nature of learners' interaction with the materials.

2.5.2.3 Hwu

Hwu (2003) explored learners' behaviour through tracking technologies in computer based activities in order to ascertain whether students actually do what the instructor intends them to do in CALL. WebCT's tracking system was used to study learners' program access behaviour. The course comprised 19 lessons and each lesson used video based MM input activities with soap opera clips to teach two Spanish past tense forms. As the students worked on the materials, WebCT kept track of the page name, the path the students took to go from

one page to another, and the time a particular page was accessed. A computer program was developed to study the duration of the learner's stay on each page and to read the tracking data, and reports on the students' access patterns and time usage were generated. The students' access to course materials outside class time was also recorded.

In real instructional environments variables such as lost connections, leaving the terminal, getting interrupted and printing affect student tracking data. For example, Hwu identifies the following variables: (1) a student can leave a window open and click another page, and while answering the questions the student can play the video again, making the duration shown on the video page less than the time the student actually spent watching the video. (2) The student may play but not pay attention to the video clip. He or she could be talking to a fellow student so that the time recorded for that page would actually be longer than the time spent on the page. (3) A student could print out a page to read so that more time was spent on reading the page than was actually recorded.

On account of these variables, Hwu decided not to focus on how long the students spent on each page but just differentiated between very short, short and slightly longer periods of time spent on activities, and whether a page was accessed. Duration of less than 10 seconds (just a click), between 10 and 35 seconds and more than 35 seconds are all given the value of 1 and studied in comparison with each other. How measuring a less than 10 second click is meaningful is not very clear unless as a comparison point only. The results showed that most students spent more than 35 seconds on video clips and clicked through the explanation pages. A few pages were not viewed at all by a few students.

Hwu administered a single 'survey' question to students at the end of the course about whether these web based lessons improved their use of past perfect forms. The majority of the students agreed that the course helped them. To evaluate the effectiveness of this treatment a *t*-test was used between the pre-test and post-test results of the control and treatment groups. $P=.0026$ (p value $<.05$), indicating that treatment had an effect. Pearson correlation analyses showed that students' self evaluations of their knowledge gain had a high correlation with their test scores and also that the more explanation pages they read the higher the increases in their scores. The findings also suggested that the students who accessed explanation pages less had lower increases in their scores.

The statistical results of this study are based on a small number of samples and therefore should be studied with caution. However, they are supported by the observation of the students' behaviour in class by the instructor and tracking data prior to the computation of statistical results. Hwu suggests that this method has implications for instructors because it can help them introduce a new classroom model where the computer is used to deliver knowledge, learners work independently according to individual learning differences, and tracking technology monitors behaviour, leaving the instructor free to observe and intervene as a coach only when needed.

2.5.2.4 *Beatty and Nunan*

Beatty and Nunan (2004) sought to evaluate the impact of MM CALL materials on learners by engaging in in-use evaluation. They investigated how language can be learned incidentally through collaborative learning in an experiment where they tried to determine whether a constructivist model of design interface generates greater collaboration than a behaviourist model. Two interfaces were designed using Mary Shelley's novel *Frankenstein*, (CD-ROM *Frankenstein Illuminated*); one followed a behaviourist model of instruction with tasks and hints and the other featured a constructivist model of instruction in which the same tasks and hints were organised in a game, set in a graveyard with horror sound effects and other multimedia features. Students were required to work on seven tasks in pairs and dyads were randomly assigned to behaviourist or constructivist versions of the materials.

Data was gathered using video recording and *Lotus ScreenCam*. The Lotus ScreenCam software captured mouse moves and keystrokes thereby enabling tracking of the dyad's navigation through various parts of the interface. Ten hours of video were transcribed but only one or more episodes from each of the ten sessions were selected for further analysis. The data was analysed both quantitatively and qualitatively. Beatty and Nunan (2004) distinguished between behaviourist and constructivist models on the basis of "locus of control" and suggested that the control resides with the learner in the constructivist model and with the teacher in the behaviourist. It is not clear however, how the determining of navigation patterns through keystroke logging reveals that collaboration is taking place. Data from Lotus Screen Cam would only become meaningful in this case when analysed alongside data from video recordings. Beatty and Nunan (2004) concluded that their hypothesis of a

constructivist interface yielding more collaboration was not upheld and give a convincing explanation as to why this may have been so without critiquing their research design or the methods of data elicitation.

2.5.2.5 Smidt and Hegelheimer

Smidt and Hegelheimer (2004) aimed to investigate how ESL online teaching can be informed by authentic web delivered video improving the incidental acquisition of vocabulary and listening comprehension. Their study explored academic listening, incidental vocabulary acquisition, learner strategy use, task effectiveness in listening comprehension and learner task interaction. The performance measures used were partial dictation tests: pre-tests, post-tests and delayed post-tests, questionnaire, retrospective interview, along with a CALL activity featuring an academic lecture. They also used Camtasia Recorder (2001) (a screen capturing software) to record interaction during the CALL activity (audio of the lecture and any utterances from students) which was also later used for a post task interview. The procedure followed was that 24 students completed the partial dictation tests and the CALL activity. Based on the pre-test partial dictation scores, the online interaction with the CALL activity of nine participants was captured using Camtasia Recorder. The same nine participants were then invited to participate in a retrospective interview intended to shed light on learning strategy use. The research method used in this study was a “within-subject design” using repeated-measures ANOVA for analysing performance data. Post-task interviews were qualitatively analysed in the light of usage data from the Camtasia Recorder in the form of screenshots of learner task interaction. As compared to the Beatty and Nunan’s (2004) study this use of screen capturing software was more effective and revealed more meaningful data because it was used in conjunction with post-task interviews and could be triangulated with the (limited) statistical data from the test scores.

On the basis of this study, Smidt and Hegelheimer (2004) recommend the use of technologically advanced screen capturing applications to track learners’ on screen behaviour, with the proviso that analysis of such data is time intensive. However, although Screen capturing software may be good at recording mouse movement and verbal utterances it fails to capture off screen activities like note-taking (Cohen 1998). Smidt and Hegelheimer see a problem in that the verbal utterances recorded by Camtasia software may actually interfere with the comprehension of the listening task by interfering with the activity’s audio output. They are sceptical about the value of recording utterances which are not necessarily “indicative of cognitive processes and/or learning

strategies". Whereas I concur with their observation I also feel that they are making too gross a generalisation based only on their experience and the context of their study. For activities which integrate the skill of speaking with listening Camtasia's recording capability would be very useful. Smidt and Hegelheimer further suggest that intensive learner training prior to recording introspective verbal protocols would result in more reliable insights into cognitive processes and learner strategies. However, they do not clarify how this would address their earlier concern about verbal utterances interfering with the activity's audio output.

2.5.2.6 *Hémard and Hémard and Cushion*

A series of evaluation studies of online materials are reported in Hémard (1998, 2004) and Hémard and Cushion (2003a, 2003b). These studies stem from the authorable CALL project developed at London Guildhall University (LGU), which aimed to design and develop an online, text-based and multimedia authoring tool in order to create and integrate a web-based interactive CALL environment into language curricula. Its objectives were to develop web based materials and evaluate their usability and utility within the university's language student population.

Hémard (1998) attempted to strengthen the theoretical basis of hypermedia CALL authoring by investigating the use of mental models in HCI literature and establishing a connection between product design specifications, the analysis of user requirements, and program usability. The central concern of Hémard (1998) is not evaluation or methods but determining the theoretical principles that should influence design for greater usability and acceptability.

Since mental models are associated with the learning process, performance and user interface design, the methodology used in Hémard's study took a three pronged approach. Three methods were used to explore these three aspects of mental models: (i) task based user walkthroughs to assess performance, (ii) experimenter-led group discussions for the learning process and (iii) feedback sessions for assessing design. The user-walkthrough was based on verbal protocol analyses and cognitive walkthroughs (Polson *et al.* 1992; Wharton *et al.* 1994; also Cf. section 2.4.1.2). These walkthroughs recorded the students' intentions to perform actions, the outcome of these actions, and reflections on the evaluation of the outcome. At the end of the series of walkthroughs, a student audit was conducted to encourage students to translate their perceptions and reactions into design considerations. This audit gathered reflective evaluative data relating to the user interface design.

Evaluations were conducted for a whole year, in three stages. The first stage used verbal protocols to familiarise students with the environment and record their views. The second stage invited students to comment on their performance during specific tasks and activities, and the final stage invited students to reflect on their experience of, the program's structural dimensions, and the embedded learning strategies. This audit was conducted by using a checklist which included criteria such as visual clarity, functionality, flexibility error prevention, student guidance and support.

Although this evaluation yielded interesting results the author does not elaborate on how they were arrived at. There is no mention of how data was recorded, or whether data was transcribed and codified for analysis. The research design makes sense but the absence of any exemplars from the data and the missing detail of analysis expect a leap of faith from the reader. The paper does not report on experimenter-led group discussions or interviews, although these were part of the evaluative methodology.

Hémard and Cushion (2003a) gives an account of an evaluation of an online test facility built into the CALL project at LGU. The paper also discusses methods and their most appropriate use in relation to the design of web based materials, and looks at the use of 'mental models' to identify key specifications of a good user-centred design (2003a:121). Hémard and Cushion emphasise the notion of the 'acceptability' of the tool, by which they mean students' perceived benefit from it, through gaining a sense of empowerment and enjoyment (Hémard and Cushion 2003a:124). The notion of acceptability was reflected in their user-centred approach to design which gave priority to "...usability over functionality" (ibid: 124). They also considered performance to be an important measure of usability and utility, and linked it to acceptability because they felt that performance is mainly responsible for "generating long lasting and efficient output" (ibid: 124).

Hémard and Cushion (2003a) used a series of post-use questionnaires as their evaluation method but faced the commonly expected problem of a relatively low response rate. The rate of return in their study was 52% (15 out of 29), they gave 'student fatigue' as a reason for this (Hémard and Cushion 2003a:135). At various times their project employed questionnaires as a method both for formative (Hémard and Cushion 2003a) and summative (Hémard 1998) evaluations.

Hémard (2004) is a paper that deals directly with issues central to this thesis, namely evaluation methods, albeit in the context of “enhancing online CALL design”. It discusses user-centred evaluation methods, premising the discussion on HCI literature and identifying the advantages and disadvantages of different methods and the most appropriate time for their use in the design process. In a table, Hémard compares how and when to use *Questionnaires, Informal feedback, Real life Observation, Checklists, User walkthrough, Focus Groups, Tracking, Usability Testing and Heuristic Evaluation*, and also compares the type of data each method reveals. The vantage point is that of the materials designer, however, not the teacher evaluator seeking to select the best instructional package from existing available materials. Nevertheless this paper is significant because it draws on HCI methods in order to evaluate CALL materials, and uses evaluation principles for usability assessment from the HCI literature that should also underpin CALL materials design and evaluation.

The three methods that were implemented in the case study Hémard reports in this paper are (i) a questionnaire, (ii) user walkthroughs, and (iii) focus groups. The questionnaire used a Likert scale and was meant to elicit CALL-specific feedback. Four 2-hour user walkthroughs for three groups with two students in each group were conducted. Two 2-hour focus group meetings were held. The paper compares these three methods on the basis of input and output, the level of involvement of staff and students, and the time and resources used, but although the study yielded plentiful data and interesting findings, as in Hémard’s previous studies, the methods of data collection and analysis are not fully explained. The user walkthrough and focus group data was analysed qualitatively and inductively, but there is no way of ascertaining how intrusive these methods were, or how the context specificity of this study influenced the reliability of these methods.

Later work by the same author (Hémard 2006a, 2006b) will be discussed in the final chapter.

2.5.3 Concluding remarks

The above review shows that researchers engaged in evaluating CALL materials have used a variety of methods, some old and time tested such as questionnaires and interviews, some new such as video observation and key logging and some borrowed and adapted such as think aloud verbal protocols and retrospective and tracking technologies. There are relatively few in-use and post-use evaluations of language teaching materials and only the most recent ones (Hémard 2004, 2006a, 2006b) appear to have consulted HCI usability evaluation principles to guide their design and evaluation of materials. However, the primary focus of the reviewed studies is not on

the appropriateness of evaluation methods but on the use of research methods for evaluation purposes. In HCI these same research/evaluation methods may be used for software usability evaluation but may be called different names, and the evaluation objectives may have more to do with the technical functionality of the program.

Usability evaluation is a highly advanced field in HCI and those methods which could be meaningfully used in CALL MM materials evaluation were reviewed in an earlier section 2.4.1. In the following section a synthesis of the information on methods studied so far is attempted by presenting in tabular form methods enjoying currency in different domains of educational evaluation, research methods and HCI usability evaluation.

2.6 A synthesis

Table 2.3 is based on key texts in HCI literature (Dix *et al.*, 2004; Nielsen 1993; Preece *et al.* 1994) that deal at length with usability evaluation methods and educational research and evaluation methods literature (Bennet 2003; Bryman 2004; Creswell 2003; Cohen *et al.* 2003; Mertens 2005; Richards 2003).

Table 2.3 Methods across Disciplines: A Composite List

Composite list	Research and Evaluation Methods in Education	HCI Usability Evaluation Methods
Observation & Monitoring	Observation Structured, semi-structured, unstructured (Bryman, 2004; Bennet, 2003; Creswell 2003; Cohen <i>et al.</i> 2003; Richards 2003; Mertens 2005)	Observational techniques: Think aloud cooperative evaluation (Dix <i>et al.</i> 2004) (Nielsen 1993) Observing and monitoring usage (Preece <i>et al.</i> 1994)
Questionnaires and Interviews	Questionnaires (Bryman, 2004; Bennet, 2003; Cohen <i>et al.</i> 2003)	Query techniques (Interviews and questionnaires) (Dix <i>et al.</i> 2004)
	Interviews (Bryman 2004; Bennet 2003; Cohen <i>et al.</i> 2003; Richards 2003)	Questionnaires and interviews (Nielsen 1993) User's opinion (Preece <i>et al.</i> 1994 :615)
Empirical methods	Empirical methods pre-post tests and experiments. (Bryman,2004; Cohen <i>et al.</i> , 2003)	Empirical methods (Dix <i>et al.</i> 2004) Experiments or bench mark test

		(Preece <i>et al.</i> 1994)
User Groups (Focus Groups/ Logging actual use/ User Feedback)	Focus Groups (Bennet,2003)	Focus Groups (Nielsen 1993)
		Logging actual use (Nielsen 1993)
		User Feedback (Nielsen 1993)
Cognitive Walkthrough	Verbal protocols/ Think-aloud Protocols(Cohen <i>et al.</i> ,2003)	Cognitive Walkthrough Cohen <i>et al.</i> , 2003 (Dix <i>et al.</i> 2004; Wharton <i>et al.</i> 1994)
Heuristic methods	Checklist Based Evaluation. (Hubbard,1996; McGrath,2002)	Heuristic Evaluation (Dix <i>et al.</i> 2004)
		Heuristic evaluation (Nielsen 1993)
Review Based methods	Expert Review Evaluation /Meta Analysis Studies	Review Based Evaluation (Dix <i>et al.</i> 2004)
	Document Study (Bennet ,2003; Cohen <i>et al.</i> ,2003)	
Model Based methods	Case Study	Model based Evaluation
	(Bassey 1999; Nunan 1992; Mertens 2005)	(Dix <i>et al.</i> 2004) Prototype analysis Analytical modelling (Preece <i>et al.</i> 1994)
Predictive Evaluation	Predictive Evaluation (Ellis,1997, Squires and McDugall,1996)	Predictive Evaluation Nielsen 1993
		Predictive Evaluation (Squires and Preece 1999) Preece <i>et al.</i> 1994
Retrospective Evaluation	Retrospective Evaluation (Ellis,1997,pp36)	Interpretative Evaluation Nielsen 1993
Context based Ethnographic Inquiry	Ethnographic Accounts (Cohen <i>et al.</i> ,2003)	Contextual inquiry
		Participative evaluation
		Ethnography

In section 2.4.1, Table 2.2 lists generic methods used in HCI, and Table 2.3 includes research methods. Relating these two tables to the CALL studies reviewed in section 2.5.2 (keeping methods in focus) will help in further synthesis.

A few studies reviewed in 2.5.2 have used computer-based video capture recordings and keylogger software (*Camtasia* by Schmidt and Hegelheimer 2004, and *Lotus Screen cam* by Beatty and Nunan 2004) while non computer based video observation was used also (Legenhausen and Wolff 1990). Hwu (2003) studies navigation

patterns derived from an especially designed program to work with WebCT's activity monitoring software. Desmarais *et al.* (1998) also used 'on-line tracking' program: an auto data logging software. Activity-monitoring software captures screen shots of end-users interacting with the program while user actions are recorded by the auto-logging program. These two methods are coalesced as one in some applications. They are relatively new to the domain of CALL materials evaluation and can be instrumental in studying learner behaviour and navigation patterns through MM materials. Navigation patterns have also been reconstructed using verbalisation techniques (Desmarais *et al.* 1998). Verbal protocols and user walkthroughs as elicitation techniques for strategies have been used in quite a few studies (Desmarais *et al.* 1998; Hémard 1998, 2004; Hémard and Cushion 2003a; Legenhausen and Wolff 1990). Focus groups discussions have been used by Hémard (1998, 2004).

In this section different methods have been collated in an attempt to derive a composite list that may aid my selection of a few evaluation methods for trialling. As is evident from Tables 2.2 and 2.3 the choice of methods is vast but my approach is eclectic and both quantitative and qualitative methods can be considered. The selected methods should however be user centric/ learner centred. All the selected methods will evaluate usability, but with variations on the theme. For instance one method can be more quantitative another more qualitative, one can be highly subjective and the other completely objective. Such a scheme may also address concerns of methodological triangulation. As the particular complexity of this study makes the evaluation methods also the research methods of this study they will be revisited in the methodology chapter explaining their implementation

2.7 A selection for closer study

I will be explaining in Chapter Three how I arrived at my final choice of methods. I also reviewed the literature in the field on these methods and I decided to include it here as part of the theoretical framework because methods themselves are a focus of this study and it is appropriate that they have a place in this chapter. The literature that was reviewed identified a few methods that are employed across disciplines and proposed in most recent literature on CALL evaluation. Their most distinguishing aspect which recommends them for use in this study is their user-centeredness and their complementarities to each other if used in tandem.

2.7.1 *Focus groups*

Focus groups are organised to explore individuals' views and experiences through group interaction (Bloor *et al.* 2001; Litosseliti 2003; Puchta and Potter 2004). Focus groups can be described as a carefully structured and planned discussion with the objective of obtaining perceptions about a particular subject in a non-threatening and permissive environment where participants share ideas and perceptions and respond to each other (Bloor *et al.* 2001; Litosseliti 2003; Puchta and Potter 2004). Groups are focussed because they are involved in a collective activity around a few particular issues. They are interactive because the group dynamics and exchange of ideas are of supreme importance. The participants build on views expressed by others in a synergistic manner, which generates insightful information (Litosseliti 2003: 2; Krueger 1994: 6).

Most definitions of focus groups have the following two core elements: a trained moderator who sets the stage with prepared questions or an interview guide and the goal of eliciting participants' feelings, attitudes and perceptions about a selected topic (Bloor *et al.* 2001; Litosseliti 2003; Puchta and Potter 2004). Morgan (1997) describes the history of the focus group as falling into three periods: early work carried out by applied social scientists and academics; market research between the period of the second world war till 1980; recent academic research, market research and political settings where focus groups have been conducted in many varied fields.

The use of focus groups can be differentiated according to the desired outcome and the type of research question. Focus groups can be self-contained or can be used as an adjunct to other research methods such as individual interviewing, participant observation, surveys or experiments. According to Wilkinson (1998) the two most common research designs involve the use of focus groups in an initial exploratory hypothesis-generating phase and in a final follow-up phase that pursues the exploratory aspects of the analysis. When focus groups are used in the context of marketing either in profit making or non-profit organisations the research is mostly concerned with an evaluation of marketing products and or services (Puchta and Potter 2004: 4-8).

In the literature on software evaluation and usability testing group interviews are used in evaluation meetings but the term focus group is not used to describe them. In general educational evaluation focus groups (and more recently virtual focus groups) are used to gather data about teaching/learning programmes but in ELT materials evaluation I have not come across any studies which have used focus groups.

The advantages of using focus groups are that they present a more natural environment than individual interviews because participants are being influenced by others and are influencing them in turn just as it happens in real life (Krueger 1994:19). The interaction and stimulation among the group participants ends up unravelling new open-ended directions for discussion (ibid).

The size of the group will be dictated by logistic issues and will partly reflect the nature of the topic and the characteristics of the individuals involved in the group (Bloor *et al* 2001; Litosseliti 2003; Puchta and Potter 2004). Some researchers favour smaller groups as they may be closer to a normal setting for discussion and would allow each member to contribute more (Bloor *et al* 2001). Larger groups may leave the individual group members feeling frustrated if they have not been given sufficient time to express their views on the topic and may be harder to facilitate and result in problems during transcription and analysis (Bloor *et al* 2001).

According to the literature (Bloor *et al* 2001; Litosseliti 2003; Puchta and Potter 2004; Krueger 1994; Stewart and Shamdasani 1990) focus groups typically consist of six to ten participants. Smaller or larger groups can be accepted depending on the research purpose. However it is difficult to manage, moderate and analyse large groups (Puchta and Potter, 2004). Smaller groups are more manageable and can be used when the aim is to explore complex controversial topics and to obtain detailed accounts from the participants (Litosseliti, 2003). Furthermore, small groups provide more opportunity for people to talk (ibid).

In most focus group research there will be generally more than one focus group with different groups of people working on the same topic (Bloor *et al* 2001; Litosseliti 2003; Puchta and Potter 2004). This is because the findings of any single session may not be revealing enough and also because focus groups require several representative users (Litosseliti 2003).

The participants are generally homogeneous like-minded individuals from the same gender, ethnic, sexual, economic or cultural background (Litosseliti 2003). The length of a focus group would typically be one and a half to two hours and sessions are tape-recorded to facilitate analysis (ibid).

Puchta and Potter (2004) emphasise the key role of the moderator in facilitating interaction. A moderator guides the discussion using a number of pre-determined and carefully developed open-ended questions, but with minimal intervention (ibid). The moderator has to have good communication skills, managing skills and

interpersonal skills to keep the discussion on the track without inhibiting the flow of ideas and ensuring that all members participate and contribute to the discussion (ibid). The moderator cannot allow the discussion to become dominated by one or more loquacious or imposing personalities. Another important role of the moderator is to keep the group focused by not allowing them to digress from the key issues being discussed (Bloor *et al* 2001; Litosseliti 2003; Puchta and Potter 2004).

The moderator's presence can also influence the data gathered. (Bloor *et al* 2001; Litosseliti 2003; Morgan,1998; Puchta and Potter, 2004; Krueger, 1994; Stewart and Shamdasani 1990). Therefore to minimize bias and the risk of manipulation from the moderator and the participants careful consideration has to be given to how leading is the moderator's input and how openly participants relate to one another and to the moderator (Bloor *et al* 2001).

Focus groups should not be used on topics which the participants have no knowledge of or incomplete knowledge of and which do not encourage a multiplicity of perspectives and points of view to be expressed (Bloor *et al* 2001). Litosseliti (2003) summarises the reservations of Krueger (1994) Morgan (1993) and Gibbs (1997) as follows:

Potential limitations of focus groups:

- *Bias and manipulation:* danger of leading participants and encouraging them to respond to your own prejudices; participants saying what they think you want to hear
- *'False' consensus:* some participants with strong personalities and/or similar views may dominate the discussion, while other may remain silent
- *Difficulty in distinguishing between an individual view and a group view:* groups sometimes appear more consistent than they are because individuals who disagree may not say so; groups often generate more emotion than any of the individual participants may feel about the issue; individual behaviour is subject to group influence
- *Difficulty in making generalizations* based on the focus group information (not only because of the limited number of participants, but also due to the difficulty of having a real representative sample)
- *Difficulty of analysis and interpretation of results* (due to the open-ended nature of focus groups, and the influence of many immediate situational factors) (Litosseliti 2003:21).

Despite these reservations focus groups may prove to be an effective instrument for evaluating CALL MM software. Other than the obvious purpose of triangulation focus groups could yield data on participants' attitudes and give them an opportunity to reflect on their experience and verbalise the experience. They would

be particularly useful as a means of 'extending public participation' (Bloor 2001:17) when CALL MM materials are being introduced as an innovation.

When considering the composition of groups, the literature advises that care should be taken to avoid groups that consist of individuals too diverse to obtain a sufficient depth of information on the research topic. Groups should not contain individuals with conflicting views although pre-existing social groups may be used. For example; a support group for those suffering with depression or in the case of this research a group studying at a pre-sessional or another skills related course. Pre-existing groups have the advantage of providing a more natural setting for discussion and are easier to recruit. Groups of strangers can also be used for focus groups and may be advantageous where the researcher is concerned with 'over-disclosure' which in pre-existing groups might have repercussions once the research is over (Bloor *et al* 2001).

Bloor *et al* (2001) opine that ensuring that individuals attend the focus group is a particular problem for the focus group researcher and it is standard practice to recruit more participants than you actually need in the assumption that a number will not turn up on the day. They add that attendance is likely to be higher if the group consists of a pre-existing social group. Utilizing an established meeting venue and time for a formal pre-existing group can also improve attendance (ibid). Despite the fact that the researcher may face difficulties ensuring attendance yet individuals may be more likely to attend a group than a one-to-one interview as they may feel reassured that they are a part of a group of individuals who share a particular characteristic or experience and that attention is on the group rather than the individual (Bloor *et al* 2001). This could be particularly relevant to the context of this study as the participants are students and the researcher on occasion is also their teacher, and they feel bolder expressing an opinion as a group rather than as individuals. They would gain confidence from each other.

2.7.2 *Retrospective protocols*

Think-aloud protocols (TAP) have been used extensively in psychology and cognitive science to understand the mechanism of human behaviour and cognition. Ericsson and Simon's (1993) work provides a theoretical framework for TAP experiments. According to their understanding, human cognition is the processing of information which is being received by the mind directly or is being retrieved from memory. Information is kept in different memory stores, with varying access and storage capabilities: whereas short-term memory (STM) is

characterised by easy access and severely limited storage capacity, long-term memory (LTM) is characterised by more difficult access and larger storage capacity. Only information present in STM, that is, information which is being heeded by the subject (static and conscious “knowledge states” rather than dynamic and unconscious cognitive processes), can be directly accessed and reported. Ericsson and Simon’s (1993) discussion can be summarized as follows:

- Concurrent verbalisation or thinking aloud, provides data on the mental states cognitively noticed or heeded by individuals carrying out a task.
- These mental states then lead possible information about the relevant mental processes that are involved while the task is being performed.
- Under the right circumstances where there is minimal external interference (verbally encoded information, no social interaction, no interferences, no instruction to *analyse* thoughts), verbalizing is assumed not to interfere with the mental processes and to provide a faithful account of the mental states occurring between them.
- The generalizability and the relevance of the data obtained through think-aloud a protocol is difficult to assess.
- Think aloud protocols could be akin to maintaining a journal during an activity but less time consuming and intrusive and they are audio or video recorded.

Færch and Kasper (1987) apply Ericsson and Simon’s model to second language (SL) learning and suggest that it is a “sine qua non” that information which is being verbalised in introspective methods whether concurrently or retrospectively is being “processed under the informant’s attention” (Færch and Kasper 1987: 2). They therefore conclude that problem solving processes in SL would lend themselves better to this method as opposed to other processes which are activated in SL learning (ibid :2).

Think aloud protocol, as they are used in the evaluation of software usability, entails users working with an interface (GUI graphic user interface) and being encouraged to think aloud or make overt the mental processes that they may be going through at each moment. (Preece *et al* 1994:623). They are voicing what they are thinking and wondering. This information is being recorded both electronically and by the observer.

Seliger and Shohamy (1989:169) describe the different categories of verbal reporting protocols as follows:

Verbal reporting refers to a set of data collection techniques where subjects report orally to the researcher on the processes they are engaged in while performing a cognitive or linguistic task (Cohen and Hosenfeld 1981, Mann 1983).

Seliger and Shohamy (1989) state that verbal reporting has been used in research in SLA studies and cite Brown's (1981) and Brown *et al's* (1983) work on writing of summaries of reading texts Mann's (1983) work on collecting self-reports of readers identification of their reading problems and their eradication. They also cite Olshavsky (1977) and Færch and Kasper (1987) who worked on reading and collecting verbal reports from readers. Seliger and Shohamy (1989:169) identify three main techniques for eliciting verbal reports. Firstly, think- aloud is explained as making overt the thinking while working on a given task irrespective of how trivial whatever occurs to them. Secondly, introspection is explained as requiring the subjects to observe the mind processes and report on them as they occur while they are working on a particular task. Thirdly, retrospection requires the subjects to comment and give information after completion of the task. This involves subjects remembering what they experienced and the strategies they employed during a particular mental event under observation (ibid). Retrospection involves information being retrieved from long term memory as opposed to the reliance on short term memory in the first two methods (Cohen and Hosenfeld 1981; Ericsson and Simon 1980; Færch and Kasper 1987; Garner 1987)

Another explanation could be based on the temporal distance between action and verbalisation on the mind processes or evaluative comment on the action or the written account of these thoughts. This has been explained as "recency effect" by Færch and Kasper and is important to the validity of the data by (1987:15). They explain further, "It can be viewed as an open continuum, at one pole of which verbalisation coincides with cognitive activity, the distance between them increasing towards the other pole" (Færch and Kasper 1987:15). Introspective data can be placed on this continuum and distinguished as "(1) simultaneous; (2) immediately consecutive; and (3) delayed consecutive introspection." (Ibid: 15). Introspection categories (2) and (3) are also described as immediate and delayed retrospection and can be used in most language use situations whereas (1) lends itself to all SL skills translation and written task taking but not speaking (Færch and Kasper 1987:15).

Diary studies and other means of recording learners' experiences over a period of time are examples of delayed retrospection (Færch and Kasper 1987:15).

Think aloud protocols/introspective /retrospective methods are used in software evaluation because they focus on the problems a user has when interacting with a particular interface, hence this method's efficacy in debugging software (Preece *et al.* 1994:625). If the user is able to navigate the software when there are no hitches and difficulties direct observation and TAP would be of little use in the context of usability evaluation (*ibid*). In the case of language learning materials the learning task could be problematic and hence the learner would slow down. Therefore both in the case of a problem with the task or the problem with functionality the learner would slow down to think of the problem. If the navigation of the software/ learning of the language task is running smoothly and there are no problems then the TAP would not be able to keep pace with the mind (Ericsson and Simon, 1985; Preece *et al.* 1994:622). Preece *et al.* (1994) assert from their understanding of cognitive psychology and Ericsson and Simon's (1985) work that paying attention to two things, doing the task and verbalising about it, places a strain on users and they can't maintain it for longer than a few minutes. It would be only when the mind slows down when it encounters a problem would it start verbalising. It is at this point that TAP becomes very useful for it allows the observer to make the connections between what the user is saying and doing. The observer is recording the statements that the user is making and the interactions with the interface for analysis later (*ibid*). Therefore this method would be particularly useful in formative evaluations of software (*ibid*). Since the design development and evaluation of software are an iterative process such evaluation procedures could be used both formatively and summatively (*ibid*). The usefulness of this method for designers would be more if this method is used formatively but for evaluators trying to gauge students' interaction with the method and the effectiveness of the innovation it could be used quite effectively summatively.

2.7.3 Questionnaires

Brown (1988), Cohen *et al.* (2003), Dix *et al.* (2004), Preece *et al.* (1994), and Seliger and Shohamy (1989), discuss questionnaires as data gathering instruments for research and this discussion of attributes can be extended to questionnaires as evaluation instruments. In language research questionnaires are used to collect data on phenomena that are not easily observed such as student attitudes and motivation (Seliger and Shohamy

1989:172). They are also used to obtain background information about research subjects such as age, previous background in language learning, number of languages spoken and years of studying the language (ibid:172).

Questionnaires have a number of advantages. According to Seliger and Shohamy (1989) questionnaires can be given to a large group of subjects at the same time therefore they are less expensive to administer than other procedures such as interviews. They also suggest that because in most questionnaires the respondent's identity is kept anonymous they tend to share information more openly and readily. Moreover, since the same questionnaire is given to all the respondents the data collected is more uniform and standard. The fourth advantage that they mention is that since the questionnaire is given to the subjects at exactly the same time therefore the data is more accurate.

The disadvantages of questionnaires are that they require considerable care and time to develop and have to undergo many stages of refinement (Wilson and McLean 1994). Moreover the data gathered may be limited in sophistication and scope as the respondents may not have answered questions with due care. Data could also be limited due to insufficient flexibility of response in certain kinds of questionnaires (ibid: 3).

There are three main types of questionnaires structured, semi-structured and un-structured questionnaires and determining which kind should be used for a given research depends on the size of the sample. Cohen *et al* (2003:247) suggest that if size of the sample is large more structured, close and numerical questionnaires will work better; for a smaller sample size, less structured, more open and word-based questionnaire may be appropriate (Cohen *et al* 2003). Highly structured questionnaires give results suitable for statistical treatment and analysis. When the objective is to measure, then the more numerical quantitative approach is best, on the other hand where rich and personal data are needed than a word-based qualitative approach works better (Cohen *et al* 2003:248). Between the two categories of highly-structured and loosely structured lies the semi-structured questionnaire. In the semi-structured questionnaire there is a clear structure and the series of questions, statements or items are presented in sequence and the respondent is asked to answer, respond or comment on them in a way that she or he thinks best.

Cohen *et al* (2003) suggest that different kinds of questions can be used in questionnaires, multiple choice questions, dichotomous questions, rank ordering and rating scales and open-ended questions. Dichotomous

questions would feature on a more highly structured questionnaire and would ask for a yes no answer. These questions are useful because it forces the subjects to take a position on an issue. They also make coding responses easy, as they are only two categories of response. Dichotomous questions have to be used carefully because there may be very few complex or subtle questions, which can be answered with a simple yes, or no. A yes or no may be inappropriate in a situation, which is complex and a series of questions may be needed.

In order to answer questions about complex issues multiple choice questions can be used because they provide a list of likely responses to given statements. The categories in a multiple choice questions are discrete having no overlap and being mutually exclusive. Like dichotomous questions multiple choice questions can be quickly coded and aggregated to give frequency of response (Cohen *et al.* 2003: 250).

Another type of questions identified by Cohen *et al.* (2003) is rank order questions. Rank order questions are similar to multiple-choice questions in that they provide options to respondents to choose from but move beyond multiple-choice items in that they asks subjects to identify priorities. This enables a relative degree of preference, priority, and intensity to be calculated. In rank ordering a list of factors is given and the subject is asked to place them in a rank order. Rankings are useful in showing degrees of response and are similar to rating scales.

Rating scales are of many different kinds, namely Likert scales, semantic differential scales, Thurstone scales, Guttman scale. Rating scales are very useful devices for the researcher as they help in identifying degrees of sensitivity and differentiation of response while still generating numbers. In this sense they generate both quantitative and qualitative data. Rating scales are more sensitive instruments than dichotomous scales. Yet they are limited because respondents have to choose from a given set of choices (Cohen *et al.* 2003:253).

Cohen *et al.* (2003) further elaborate that a questionnaire can be made more respondent friendly by including open-ended questions to which respondents can reply in their own terms and keep their own opinions. They opine that open-ended question is ideal for small scale research or for those parts of a questionnaire that ask for a personal comment from the respondents along with ticking numbers and boxes. Open-ended responses provide useful information that might have escaped the other closed questions and also puts the responsibility

and ownership of the data more securely with the respondent. Open-ended questions yield responses that are authentic, rich deep, honest and candid (Cohen et. al, 2003).

2.7.4 *Activity monitoring and keystroke logging*

Activity monitoring and keystroke logging software as a research tool in language teaching and SLA has been used primarily to study pausing behaviour and pause-location in writing (Spellman Miller 1996, 2002; Warren 1996). This approach involves the unobtrusive logging of the use of keys and cursor as writers compose their sentences. Software logging tools combine the use of time-stamped key presses which record the exact key and time that was used by the user with capturing in video or stills the interaction (Preece *et al* 1994: 626-27). The monitoring software used for this study captured the screen shots at various intervals and records keystrokes. This software, working stealthily in the background, records all operations made in real time and stores records of all activity electronically in log files for later analysis. The output of the log file is a highly detailed record of the activity making it possible to have explicit data on the temporal features of writing on line and the sequence of processes. In case of writing these processes are the planning, revising, generating and formulating texts.

As a software evaluation tool logging software trace the navigation routes users take, how long they pause at a particular feature, movement back and forth through the software and record them (Nielsen 1993: 216-20). The patterns of engagement that emerge from activity monitoring show how users perform their actual work and because it is easy to automatically collect data from a large number of users working under different circumstances (Preece *et al* 1994: 626). A log will contain statistics about the frequency with which each user has used each feature of the program and the frequency with which various events of interest (such as error messages) have occurred (Nielsen 1993:216-20). These statistics which show the frequency of use of certain commands and other features of the system can be used to optimise frequently used features and to identify the less frequently used features (*ibid*). These statistics can identify the frequency of various error situations and mark the use of online help thus can prove helpful in improving the usability of future releases of the system by redesigning the features causing the most errors and most access for online help (Nielsen 1993).

Nielsen (1993) explains that the procedure of collecting logging data usually involves instrumenting low-level parts of the system software, such as keyboard and mouse drivers, or by modifying the software of interest. He further explains that modifying the software is better because it makes it easier to log events of interest as just

obtaining data as raw input and output makes it much harder to analyze the higher-level or more sophisticated events of interest for system usability, such as use of certain features or error situations (ibid 216-220). Also of interest to software developers would be the possibility of being able to log complete transcripts of user sessions for analysis of patterns of use, such as what commands are issued next after an error situation (Nielsen 1993: 216-220; Preece *et al* 1994:627).

Yet another use of logging data identified by Nielsen (1993) is studying the user's detailed use of a user interface to find usability problems that may not be apparent when just observing users. He explains that logging data can clearly show the statistics of one action (e.g. clicking on an object) from many users by integrating the logging data with the user interface (Nielsen 1993; Preece *et al* 1994:627). Since the logging data only shows what the users did but not why they did it is considering prudent to combine logging with other methods such as interviews, where users are shown data about their own use of the system and asked to explain their activities (simulated recall interview/ focus group interviews or think aloud protocols), (Nielsen 1993; Preece *et al* 1994:628).

In essence activity monitoring and keystroke logging is combining the functions of a cognitive walk through and observation except it is being done electronically and unobtrusively with relatively no input from the observer in its signified role in a walk through.

2.8 Summary

In this chapter the central concerns of the topic of evaluation that impinge on the research questions of this study have been identified and expert opinion discussed. The first major section dealt with the trends and issues of evaluation in the general discipline of education. From this general perspective the discussion moves to the specific of materials evaluation in ELT. The next section attempts to include HCI usability evaluation literature into the equation. After giving this background the review focuses on evaluation of CALL evaluation principles and gives a review of recent CALL studies with a focus on the evaluation or data elicitation methods used in these studies. This review paves the way for an attempt at a synthesis of methods from HCI usability evaluation literature and methods of educational research and evaluation used in CALL. The last section gives a detailed review of the methods chosen for closer scrutiny in this study.

This chapter has reviewed literature from different domains in the hope of identifying a niche for this study. A *Venn diagram* may help to illustrate what different domains this study is attempting to bridge.

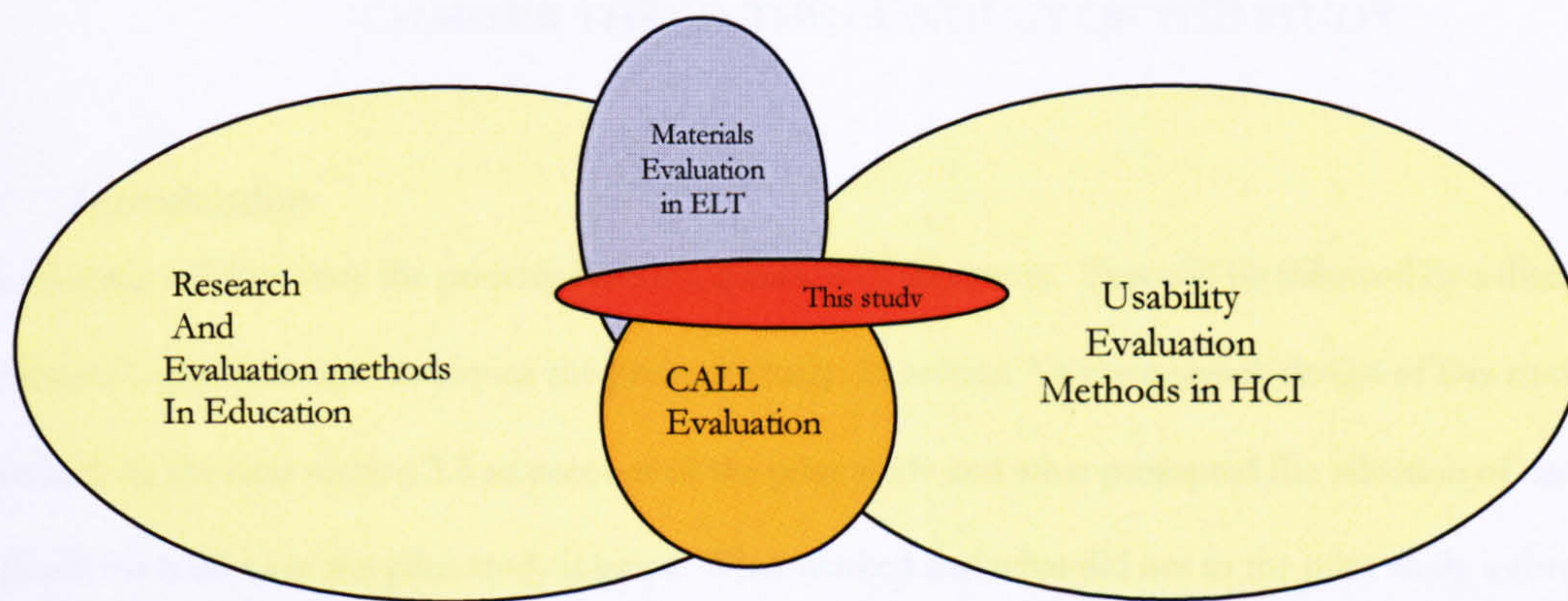


Figure: 2.6 A Venn diagram illustrating the positioning of this study

The positioning of this study is inclined towards HCI literature so that best practice in CALL evaluation can be informed by development in this highly specialised and fast developing field.

The next chapter gives an account of the conduct of the study and will include a description of the pilot study and the process of implementation of selected methods in the main study. It also discusses the criteria of evaluation derived from CALL and HCI evaluation literature for the assessment of the methods.

CHAPTER THREE: THE CONDUCT OF THE STUDY

3.1 Introduction

This chapter will first state the primary research question of this study. This will be followed by a discussion of the research approach that underpins the present study. In section 3.4 the research design of this study is explained. In the next section 3.5 an account of the pilot study and what prompted the selection of certain methods for trialling in the pilot study is given. What worked and what did not in the pilot study informed the modification of the evaluation methods to be used in the main study. In section 3.6 the conduct of the main study is described. My criteria for evaluating the effectiveness of methods are described in the final section of this chapter

3.2 The research question

When contemplating the choice of research methodology, Silverman advises that the most important factor to be taken into account is 'what you are trying to find out' (2001:25). This takes the form of the 'research question' and for the purpose of this thesis it is:

What is best practice in the evaluation of MM CALL materials and which methods or combination of methods can be employed for effective learner-centred evaluation?

Teachers attempting to teach in a technology-enhanced learning environment need to know how to evaluate MM CALL software and how best to exploit programs for instructional purposes. Teachers attempting to integrate new technologies into their teaching practice rely on evaluations to help them make instructional decisions. This study aims to present key features of various methods of evaluating software that yield such useful information. As the previous chapter showed there are many options available, but the decision regarding which to use can be problematic. Will a marriage between software usability testing methods, and methods used in educational evaluation, perhaps including language teaching materials evaluation methods, be possible? If so, what would such a marriage achieve?

3.3 Research paradigm, strategy and approach

This study is seeking to understand how CALL MM programs are best evaluated keeping the learners' perspective in view. With this in mind, a scientific approach that focussed only upon observable phenomena was not thought to be sufficient. A review of evaluation literature convinced me that central to effective MM evaluation is the learners' perception of the program's effectiveness and their experience and interaction with the program. This includes both observable phenomena such as how they operate the system and non-observable phenomena such as what they think of the program while working with it and after having completed working with it.

This study can be defined as a mixed-method case study. Such an approach draws on a range of research strategies and techniques and generates both qualitative and quantitative data (Mertens 2005). A multi-method approach to evaluation allows study and analysis of both the results and the processes associated with a new program and can yield rich findings generating multiple sources of data, which provide checks on the validity and the trustworthiness of the findings (Bennett 2003). This verification through multiple perspectives and idiographic interpretation which focuses on the process are key distinguishing characteristics of qualitative inquiry (Bryman 2004; Creswell 2003).

A multi method approach also addresses concerns of triangulation. According to Cohen *et al.* (2003:112), in social sciences, triangulation "attempt[s] to map out, or explain more fully, the richness and complexity of human behaviour by studying it from more than one standpoint, by making use of both quantitative and qualitative data". In qualitative research triangulation aims to enhance the credibility and validity of the results complying with the need for 'verification' (Creswell 2003). Altrichter *et al.* say that it gives a "detailed and balanced picture of the situation" (1996: 117). In this study "methodological triangulation" (Cohen *et al.* 2003:112) has been used to study evaluation methods from different perspectives.

3.4 The research design

According to the perused literature on research design the first step in the research design cycle starts with the researcher consulting secondary sources based on which a conceptual framework of what is being investigated is designed and then this may be implemented in the pilot. Based on the findings of the pilot study a research

agenda is created or refined. The next stage is the data collection stage, which is followed or accompanied by analysis and interpretation. I will now explain my research design.

The principal aim of this study is to analyse the effectiveness of methods used to evaluate CALL MM materials. In order to achieve this objective evaluation methods had to be operationalised to study their performance in-use in the most natural settings i.e. the classroom or the multimedia room (Creswell 2003). The research design of this study thus entailed conducting an actual evaluation using CALL MM materials and gathering in-use and post-use evaluation data. A review of literature was necessary in order to be able to select methods for trial. This was done by studying evaluation literature from the domains of educational evaluation, research methods in education, ELT materials evaluation, CALL evaluation and HCI usability evaluation. The overview of evaluation methods in the previous chapter was undertaken to ensure that the *construct validity* of the research design (which Nunan (1992:80) quoting Yin (1984) describes as “establishing correct operational measures for the concepts being studied”) is visibly upheld.

In order that the chosen methods could be evaluated, sets of respondents were engaged in working through the materials (EASE CD-ROMs *Listening to Lectures* and *Seminar Skills*), first in a pilot study and then in the main study. The methods were trialled in the pilot and the research strategy was refined and modified for the main study, based on the findings of the pilot. The nature of data revealed and the performance of the methods were assessed on select criteria. Two of the methods, retrospective protocols and activity monitoring by use of software provided close continuous in-use evaluation data. Post-use evaluation was conducted through two focus group interviews and usability and post-use questionnaires.

3.5 The pilot study

3.5.1 *Methods for trial*

The review of evaluation and software usability literature threw up multiple possibilities for an evaluation design. The choice of methods for this study was guided by the principles that the methods should be user-centric and should elicit a range of different kinds of data: subjective and objective, qualitative and quantitative, intrinsic and extrinsic. Retrospective protocols (the think-aloud verbal protocols and user walkthrough in the pilot study) and focus groups were chosen as qualitative methods capable potentially of producing data which could be analysed qualitatively in order to observe if the pedagogical intention of the designers was being upheld in actual practice

and what the experience was like for the students. Activity monitoring and questionnaires were chosen as quantitative methods. Yielding quantitative data, these two methods would also evaluate from the user's perspective. Activity monitoring was also intended to provide information about how the software was actually used whereas the other methods elicited what learners had to say about the program.

Verbal protocols: Two versions of verbal protocols were tried out in the pilot. The think-aloud protocol and user walkthrough.

Think aloud protocols: The users verbalised their impressions while they worked on the materials and reported on what they found easy or difficult and their opinion of the exercises. This information was tape recorded. (Cf. Chapter Two section 2.7.2 for more detail about the description of method).

User walkthrough: This method is a variation of cognitive walkthrough and pluralistic walkthrough (methods used in HCI described in Chapter Two section 2.4.1.2) and uses the recording method of verbal protocols. Two users are paired together in order to elicit information in the form of a dialogue. This is theoretically close to the cognitive walkthrough except that the users are in pairs and the expert user or developer (or the cognitive psychologist) is not part of the pair and both users are first time users of the software being evaluated. The objective was also to use a method routinely used to evaluate software in the field of software design and development and apply it to the EASE materials.

Focus group interviews were selected for gathering subjective opinions on the experience of using the software and gathering more qualitative data. The findings from focus groups would be triangulated with the findings gathered from other methods.

Observation: activity monitoring and keystroke logging were used to gather objective data relating to patterns of usage, time spent on each activity and learner choice. Paper and pen observation was undertaken initially but abandoned as the activity monitoring was serving the same purpose.

Pre and post questionnaires and the usability questionnaires: PLUM (Programme on Learner Use of Media) a questionnaire designed for evaluating instructional effectiveness of multimedia (Cf: 3.7.5.1 for more explanation) and SUMI (Software Usability Measurement Inventory) a questionnaire for measuring usability of software (

including instructional software) in the field of HCI (Cf: 3.7.5.2 for more explanation) were used to evaluate learning satisfaction, effectiveness of the user interface and technical specifications of the software. Data from the usability questionnaire could be statistically analysed.

3.5.2 *The procedure*

This study was conducted during the University of Warwick Pre-Sessional Programme for English Language and Study Skills. The majority of the participants on the course were students who had completed 16 years of education in their home country. Their knowledge of English varied but most scored below a 6.5 in IELTS. The participants of this research were volunteers with similar IELTS score of 5.5 or less. Their incentive to participate was the prospect of improved proficiency in the language and extra tuition that they would receive from me as the researcher. The twelve students who volunteered for the research were mostly from the WMG (Engineering) department except for three students from other faculties (Pure Sciences, Social Sciences and Humanities).

The students worked on EASE materials for seven weeks, working once a week for two hours. They were paired to work in tandem with their partners in a user walkthrough. Cognitive user walkthrough was also tried with one student with the researcher acting as the expert of the pair. Think aloud was explained to the students and the students working alone on their PCs were encouraged to record their thoughts self prompted. WAV recording on the PC was attempted but backup cassette tape recorders were also made available to record the protocols. Activity monitoring and key logger spy software from ACTMON¹ was downloaded. Interaction with the EASE materials was recorded on the activity monitor and key logger and regular reports were sent by the spy ware program to my email address. I also physically observed the interaction in an unstructured observation format and kept notes in an observation diary. PLUM pre-use and post-use questionnaires and the SUMI usability questionnaire were administered to these students at the beginning and end of the pilot study. The students also participated in a focus group.

¹ <http://www.actmon.com/>

Data from the protocols and focus groups was transcribed and analysed qualitatively. Questionnaire data was analysed quantitatively. Tracking data was studied and impressionistic conclusions were drawn about the efficacy of the method.

3.5.3 Findings and discussion

The methods trialled in the pilot study were assessed on the basic criteria of (a) cost effectiveness (b) ease of use and (c) nature of information. One significant lesson from the pilot study was the need for the evaluation criteria to be more detailed and rigorous. The nature of information revealed using these methods was described as either '*intrinsic*' or '*extrinsic*'. Evaluative opinion related to learners' attitudes and feelings about the effectiveness of the pedagogical content of the materials was termed *intrinsic* whereas information was considered *extrinsic* when interaction with the functionality and design was discussed. Intrinsic information is of a personal nature and is often analytical and conceptual in nature (thoughts and feelings of learners) and extrinsic is at the surface level (concerning the operation of the materials, ease of use and layout).

Focus groups as a method fared well on the criteria of (a) cost effectiveness, (b) ease of use, and (c) nature of information) despite a weakness in the nature of information in that it tended to be on occasions repetitive. An excerpt which highlights emergent issues (fluency, cultural conditioning and collectivist thinking,) from a focus group is given below. Two participants(names anonymised) are discussing the functionality of watching the video clip and answering the questions simultaneously.

Moddy: I think that if the conversation with the picture before when the person...when the 2 persons are speaking it must give... the picture and speaking must come before the question. First of all we listen for the conversation between the two and as we finish the question will come. In the case of EASE the question and then we have to click on the video clip to see the video clip and it will cover the question so we need to see the whole video clip before and then we make a click to see the question. I think that would be better because we would be viewing what we listen. But if we read the question first and then see the video clip that would not help us to improve our skills in listening.

Wicki: Yes I like working with CD based materials. My friend Moddy is right that we cannot see the questions when we are watching the clip.

In the excerpt the first student conveys his views in a longwinded fashion but he is a fluent speaker within the group. The second student is agreeing with the first quite unconditionally. This phenomenon was observed frequently throughout the interaction. Could it possibly be attributed to collectivist thinking cultures of the Far East? (Hofstede 1980, 1983).

Focus group and user walkthroughs yielded similar data, qualitative and intrinsic in nature. The difference was that the commentary on experience captured in user walkthroughs and the responses verbalised from short term memory (STM) were spontaneous. Whereas in the focus groups, students were commenting on the same issues as those which rose in user walk through s, but they were retrieved from long-term memory (LTM). In fact, focus group participants discussed the same features and issues as in user walk through despite the fact they had had time to reflect on what they had observed and experienced. This suggested that once participants had recognised or identified an issue as part of a cognitive process (in which they became conscious of it or dwelled on it), it entered LTM from STM. In other words the act of identifying and commenting on an issue fixed it in the LTM of the participants and this observation is retrievable at a later point in the focus group. There is value in this as this ensures that participants have an opinion which they can present or defend, and in a sense are prepared to debate the merits and demerits of the materials without wasting any time in formulating ideas during the focus group. It makes for livelier interaction between participants, with more time spent on convincing the others.

The user walk through in this study oscillated between a monological think-aloud protocol and a dialogical user walk through. The subjects mostly agreed with each other and built on what one partner said, rather than being critical of each other's points of view. The fact that they were in agreement augured well for the more interactive nature of the focus group. Participants started out with agreement and the more articulate members suggested more issues, which were discussed by the rest. Common background and shared experience of working on EASE introduced a degree of comfort amongst the focus group participants. These two methods strengthened and validated each other's findings, yielding data of an intrinsic nature.

Activity monitoring key loggers yielded more extrinsic data as compared to the intrinsic nature of the data from focus group and user walk through. A closer look at the data supported the findings of the intrinsic methods, for example if a user was stuck at a particular exercise or was taking too long to finish an exercise he/ she

sometimes commented on it in the user walk through and this was also indicated in both the observation notes (when they were still being made) and the pattern of usage provided by the key logging activity monitor software.

Three kinds of questionnaires were used in the study, namely: structured, semi structured and open ended. The questionnaires yielded data of both an intrinsic and an extrinsic nature. The intrinsic data was related to the nature of learning and the quality of experience and the extrinsic data consisted of comments on the ease of use of the interface and the different functionalities of the software.

Of the methods that were trialled the user walkthrough and cognitive walkthrough were found to be the most problematic, as the students had to be prompted time and again to exchange views and discuss the different features they were experiencing. The students who were working alone (not as part of a pair) recorded their thoughts more readily than the ones working in pairs. Besides, since both students in the pair were working on their own PCs the amount of time they took to complete certain exercises differed, resulting in one of them feeling disturbed when the other wanted to talk. This merited a rethinking for the main study where a modified hybrid version of the user walk through and think-aloud protocols was considered more feasible.

Table: 3.1 Results of the Pilot Study

Evaluation Method	Time of implementation	Nature of data	Ease of Use	Emergent issues
Focus Groups	In use evaluation Post use evaluation.	Subjective/Intrinsic: feelings and opinions expressed	Pre planning required but still easy to administer	Participants seemed reluctant to express opinions. Perhaps on account of lack of fluency in English? or culturally conditioned behaviour?
Retrospective Protocols, Think Aloud Protocols/ User Walk Through/Verbal protocols	In use evaluation	Subjective/Intrinsic data. Students talked about problems they got stuck on, and about their frustrations	Equipment and recording software on the computer would make transcription and analysis easier as compared to using old fashioned audio cassettes.	Students had to be prompted to record which interfered with the 'purity' of the method. Use of WAV program seen to be a better option for getting digital audio recordings.
Questionnaires	Pre and post use	Objective/Extrinsic	Easy to use	How seriously and thoughtfully did the students behave when

				filling them in?
Key Logger/ Activity Monitor Data Analysis During	In use	Objective/Extrinsic	Easy to use once installed and technology issues sorted.	Technology related issues will have to be worked out. Data easy to read for an overall impression but how should it be analysed statistically?

I decided to keep focus groups as a method for the main study despite noticing the above mentioned weaknesses in the nature of information. Paper and pen observation was very time consuming and was abandoned after the first few sessions as the activity monitor seemed to be doing the same job electronically. Cognitive walkthrough or paired user walkthrough were seen to be relatively ineffective and were abandoned in favour of using retrospective protocols both verbal and written as a hybrid between a user walk through and think-aloud protocols. This was done as these methods were not cost effective. The effort required to administer them outweighed the quality and value of data. The pre and post questionnaire and usability questionnaire used in the pilot were considered cost effective and easy to use and they elicited rich data. Therefore, they were also considered replicable for the main study.

Learning from the experience of the pilot I decided to select four methods for the main study: (i) focus groups, (ii) retrospective protocols, (iii) questionnaires and (iv) activity monitoring.

3.6 Secondary research questions

The adoption of the primary research question (stated in section 3.2) and the decision to explore further four of the methods trialled in the pilot study gave rise to a number of secondary questions:

- 1.1 What are the qualities and limitations of focus group interviews for evaluating multimedia CALL materials?
- 1.2 What are the qualities and limitations of retrospective protocols for evaluating multimedia CALL materials?
- 1.3 What are the qualities and limitations of pre-use and post-use usability questionnaires for evaluating multimedia CALL materials?
- 1.4 What are the qualities and limitations of tracking data from activity monitoring and key logging for evaluating multimedia CALL materials?

3.6.1 Rationale for the research questions

It is hoped that the insights resulting from addressing these questions might enable evaluators of MM teaching materials to maximise the efficiency of their evaluation strategies on a number of levels. An understanding of how different evaluation methods work in implementation and how students interact with MM materials has the potential to provide evaluators with some guidance as to whether the purchase of expensive MM programs will result in the desired learning outcomes. Armed with this knowledge evaluators might be able to improve the effectiveness of their purchasing decisions and the management and execution of evaluations that must inform these decisions.

As discussed in earlier chapters there are a number of definitions and purposes of evaluation. Guba and Lincoln's definition "We define evaluation as a process for describing an evaluand and judging its merit and worth" (1981: 35) is applicable to the questions of this study. In an analysis of evaluation methods for MM CALL materials a description of each method and a discussion of its merits and limitations are essential. Determining the optimal method or combination of methods of evaluation begins with a re-examination of the central purposes of evaluation. Guba and Lincoln (1981) see four main purposes of evaluation: improvement of the entity; critique of the entity; adapting the entity to a particular context and certification of the entity in the new context. All four of Guba and Lincoln's above cited purposes are borne in mind to a greater or lesser degree in this study, but the 'critique of the entity' (methods of MM evaluation in a given context with a view to understanding and improving their performance) is considered paramount. In an attempt to find answers to my research questions a critique of methods is undertaken in a given context with particular materials.

The topic of Multimedia evaluation by its very nature draws upon studies which evaluate "educational technologies". The literature on the evaluation of educational technologies suggests that evaluation of multimedia materials is an under researched or understudied area. Laurillard, (1993: 240) says "Research and development projects on educational media pay quantities of hard cash for development, lip service to evaluation, and no attention to implementation". This study aims to focus on evaluation through implementation.

3.7 The main study

In this main section the context of the study and the participants will be discussed first, followed by a detailed explanation of the procedures of operationalising the selected methods in an evaluation of EASE materials.

3.7.1 *Participants*

The participants of the main study were 40 students of the Lahore University of Management Sciences (LUMS) from the freshman/sophomore years with a Computer Science, Economics, Finance and Accounting or Social Science major. All the students had to study computer science and English Communication skills as mandatory core courses. Their ages ranged from 18-22 years and they mostly hailed from Pakistan having studied mostly in schools where the medium of instruction was English. The student respondents were selected on the following basis:

- They were available for the duration of the evaluation and agreed to be part of this study
- Their levels of achievement in the Scholastic Aptitude Test (SAT) tests both verbal (literacy) section and Math (numeric) section were very similar.
- They shared a common language (Urdu) and had close similarity in educational background. The medium of their primary and secondary school education was English, with only a few exceptions
- They all had a similar level of computer literacy
- Although they had seen and used similar CD-ROMs before, they were new to language teaching CD-ROMs
- The subjects were willing to be interviewed about their experience and were willing to participate in recorded focus group interviews.

The students taking this course had similar academic backgrounds yet there existed slight differences such as coming from the “A” levels stream or the local equivalent Intermediate FA/FSc.

3.7.2 *The procedures of the study*

The students who used EASE were offered this as part of their regular Communication Skills (SS122) Course which is mandatory for all students of LUMS. The course was a three credit-units course which meant that they had to have at least thirty hours of contact time with the instructor. I as their instructor was responsible for teaching the course (SS 122 section V) and managing related issues like assessment and tutorial support. It had two 2.5-hour long sessions per week in the multimedia lab and one-hour long session face to face with the instructor. The course was conducted over a period of six weeks with the students thus getting 12 sessions of 2.5 hours duration to work on the two EASE CD-ROMs. They were encouraged to spend extra time on the CD-ROMs if they so desired. The laboratory however was booked for twice a week for 3-4 hours, to cover for the 2.5 hours of class time.). It was a two hour session at the end of a long day and students complained about the scheduling of the sessions. This late scheduling was likely to affect their behaviour.

The students were given an introduction to EASE and were told about the nature of this research and its potential benefits to them. Ethical angles such as confidentiality, anonymity and permissions were cleared. Ethical considerations addressed in the recommendations on good practice in Applied Linguistics by the British Association of Applied Linguistics (BAAL, 1994) were kept in view. The evaluation procedures and the students' roles were explained and demonstrated. Students were informed about the monitoring of their activity and they were given the option of transfer to another section of their choice, if they wished. (During the course of the evaluation they seemed to forget that the monitoring software was working stealthily in the background which served the purposes of this evaluation well.) They were given a pre-treatment questionnaire at the outset which was collected and then given back to them at the end, when they filled in the post treatment questionnaire.

The students were also informed about the audio and video recording of their session and the fact that I would be observing their interaction by going around the laboratory. The students were allocated workstations and were told that they must use the same PC every time they logged on to work on EASE. Each student name and ID and the PC number they were working on were recorded. This helped in maintaining a record of the activity monitoring files.

At the start of the course the number of students on the course was higher (50) but after a week or two of classes and laboratory sessions the numbers dropped. The attrition rate could be attributed to students' initial curiosity disappearing - they wanted to see how this e-learning based course would be administered but then lost interest. Some may have feared that they would not do well enough in this innovative method of teaching.

The laboratory personnel of the University's ITSC department underwent some training in uploading the activity monitoring software and maintaining it for the duration of the course and evaluation. Daily screen shot recordings and key logger files from the 40 designated PCs being used by the students had to be extracted and saved. This created quite a few technological and work load issues. The support offered by the IT department was invaluable and the data could not have been collected without their expertise and willing cooperation. Data from the activity monitoring and key logger software was lifted from students PCs periodically during the course and saved on two dedicated computers from which it was transferred to compact discs at the end of course. Each laboratory session's data required 4 CDs on average. So over 40 CDs of 800MB worth of data was collected. This included screenshots taken at 10-15 seconds intervals of student working on EASE materials and key logger files of their interaction.

Retrospective protocols such as think-aloud protocols/ oral and written protocols were explained to the students in depth. An innovation of recording think-aloud protocols using WAV recording software as they worked on the software was introduced and students had to be taught how to use the microphones and the recording software. Since these students were all conversant with computers they did not experience any difficulties operating the recording other than initially in the first session. The WAV software was installed on all PCs in the laboratory. Saving the WAV recorded files on the designated space on the University's server under their student ID created a few problems initially. A few recordings were lost because of server outage (down time). The students were excited about this speaking opportunity, however, and this tied in with the speaking practice required as part of the communication course they were enrolled in. Each student made 1-2 recordings of 2 -3 minutes duration on average per session. Each student made a total of 12-15 WAV recordings. 40 students thus generated vast amounts of oral protocol recordings (480 is a conservative estimate). These recordings were saved on ten computer discs, and were easier to transcribe than the video recordings of the focus groups.

I asked the students to write me an e-mail immediately after the session, jotting down their impressions of the interaction, like in a journal entry. This was introduced to capture the impact and spontaneity of the experience adding to the in-use evaluation aspects of retrospective protocols. I hoped that these e-mails would capture the immediacy of experience which may have been lost if evaluation was delayed till the end of treatment. I expected a certain amount of thought to go into these compositions and some formality, as opposed to the informality of spoken/ oral protocols on WAV files. The comments by their very nature were meant to be evaluative and might cover some technical/theoretical aspects, not covered in the oral protocols recorded on WAV files. I received over 400 e-mails.

I administered the SUMI Usability questionnaire immediately after the treatment came to an end in the last session. The same version of the SUMI inventory, licensed to the researcher, was administered a week later. These two evaluations are referred to as Day 1 and Day 2 results when analysing the data from SUMI. Two sets of forty questionnaires (one for each of the EASE CDs) were completed by the respondents. At the end of the six week treatment I conducted two focus group interviews. They were 75 -90 minute video recorded sessions held consecutively, and students were purposively selected on the basis of their being representative of the entire population of the group that participated in the evaluation. I used a rank-order focussing exercise (Appendix 4.1) in the focus groups. Afterwards the video recording was digitised for viewing on the computer and ease of transcription.

In sum, the data collected from the four methods that had to be analysed included: (i) two 75-90 minutes of video data of focus groups; (ii) over 800 written and oral retrospective protocol accounts; (iii) data from four questionnaires on each CD-ROM (pre-use and post-use PLUM questionnaires (120 respondent questionnaires approx) and SUMI Day 1 and Day 2 results for both CD-ROMs: (160 Respondent questionnaires approx); and (iv) 45 students navigation patterns and time usage logs, taken from the activity monitoring software. Out of this data from 20 students working on different stages of the materials (units 2 and 6 from *Listening to Lectures* and units 2 and 5 from *Seminar Skills*) was selected to be prepared for analysis.

Different methods were employed for analysing both qualitative and quantitative analysis tools. QSR NVivo Version 2.0.161, a qualitative research tool, was employed to code data from the focus group interviews, retrospective protocols and qualitative data from open ended questions of the PLUM pre use and post use

questionnaires. Microsoft Excel and SPSS were used to analyse the quantitative section of the PLUM post use questionnaire. The SUMISCO analysis software designed to analyse data from SUMI was used for SUMI data.

3.7.3 *Focus groups*

A detailed description of focus groups as an evaluation method and as a data gathering instrument was provided in Chapter Two section 2.6.1. This short summary harks back to those descriptions in order to introduce an account of how focus groups were conducted in this study.

Focus groups were used for applied research in fields such as marketing long before they gained popularity as a qualitative research method in the social sciences. Focus groups have been the dominant form of qualitative data collection in marketing research for over a generation (Goldman and Macdonald 1987; Greenbaum 1993; Hayes and Tatham 1989; Mariampolski 2001). Their use in educational research is more recent (Tierney and Dilley 2001: 461). Focus groups are basically group interviews where the focus is not on the researcher and interviewer taking turns to question and respond as in typical interviews with opinion emerging through this structured turn taking. Rather the focus is on the interaction within the participants of the focus groups. The participants build on views expressed by others in a synergistic manner, which generates insightful information (Litosseliti 2003: 2; Krueger 1994: 6).

For usability evaluation of software, about 6 to 9 users are generally brought together to discuss issues relating to the program. In social sciences and marketing focus groups the number stipulated is 6 to 9 with the smaller number being used for more intense and subjective data gathering (Littoseliti 2003; Bloor *et al.* 2001).

The participants respond to selected topics pertaining to the research and a moderator presents these topics in the form of open-ended questions to the group. Considerable preparation is undertaken prior to conducting the focus group. Questions are prepared and refined by informal piloting. Moderators are trained to conduct effective, unbiased, free flowing discussion through their informed moderation. Participants in focus groups are selected as representative of the target population. Purposive sampling is thus done as opposed to random sampling, with attention given to selected criteria to uphold validity and reliability concerns (Sykes, 1990). After careful selection of participants and adequate preparation, the duration of the actual focus group interaction is an hour and a half or two hours. At the outset the purposes of research and any ethical concerns are dealt with

and the participants are told the norms of interaction (turn-taking, explaining, seeking clarifications, speaking concisely and giving specific examples to illustrate points, not speaking at the same time.) The presence of the camera is explained. This is followed by a preliminary activity, which has the purpose of helping the participants focus on the key concerns of the research or evaluation. Bloor *et al.* (2001) recommend using a rank ordering exercise as a focusing exercise prior to discussion of the main questions. The main discussion questions are put to the participants in no particular order and the questioning technique is also altered and varied from one focus group in a study to another, (Krueger and Casey 2001; Kitzinger 1995; Morgan 1997). The moderator's role is crucial to effective focus group interaction. The moderator has to be attuned to the development of discussion threads and also be able to elicit key information. After the focus group discussion has been recorded, it is transcribed and different analysis strategies may be applied. The data are categorised, coded and interpreted.

3.7.3.1 Focus group procedure

The details of the students who participated in the focus group are given in Table 3.2.

Table 3.2 Focus Groups Participant profile

Evaluation Process	Worked on CD	Participant number	Age	Majors
Focus Group 1	20-30 hours	7(6male1female)	18-21	CS/AC/SS/Econ
Focus Group 2	20-30 hours	7 (4 male 3female)	18-21	CS/AC/SS/Econ

As mentioned earlier my students were an intellectually homogenous group but coming from two streams (A-levels and FA/FSc), so care was taken that proper representation was given to both streams so that a representative purposive sample was selected. The students worked on the software for six weeks. At the end of six weeks two groups of nine students were asked to participate in the focus group interviews. Seven complied.

I played the role of the moderator of the meetings. The group was adequately prepared in terms of rules of conduct and the questions to be asked. I prepared a list of issues/ topics/questions beforehand which would assist in eliciting the relevant information from the group (Chapter Four Table 4.3). A focussing exercise using a rank-order exercise was designed to help participants focus on the topic. I developed the statements of the ranking exercise after having worked on the ASE CD-ROMs myself and determining what pedagogical objectives these materials aimed at achieving. This focussing exercise was conducted prior to the main

discussion to engage the group with the issues of evaluation and to contextualise the EASE materials. The whole sessions of the two focus groups were video recorded with the aim of capturing spontaneous user reactions and ideas that evolved in the dynamic group process, and for subsequent ease of transcription and to eliminate the risk of data loss.

On the day of the interview students all came to a room with a circular seating arrangement around a table. They were briefed on the norms of polite interaction to be followed during the filmed focus group interview. Confidentiality and ethics in research issues were explained. The nature of focus group interaction and the purpose of this research were explained to them. The students were first asked to do a round of introductions for the sake of the record and to get comfortable talking in front of the camera. They introduced themselves by giving their names and backgrounds. The ranking exercise was then explained to them (Appendix 4. 1) and they were asked to do the exercise. The first ten minutes of the interaction were spent on this exercise, which was particular to the materials that they had been working on for the past six weeks. The exercise had a series of statements pertaining to EASE and its pedagogical objectives and they had to rank these statements in order of importance after discussing them and arriving at a consensus.

As the teacher moderator I tried to keep the discussion on track and generate free flow of ideas and comments. However it was difficult to maintain an objective uninvolved stance when the discussion sidetracked or students went on the offensive about the materials. I tried to ensure that all participants got to contribute to the discussion and was also careful to guard against having a single participant's opinion dominate the discussion.

The objective was to have the discussion feel free-flowing and relatively unstructured to the participants, but to follow a pre planned script. After the focussing exercise questions about the focus, structure, organisation, functional features of the CD-ROM were posited to the group. As soon as debate on a given question was exhausted a new aspect was introduced. Participants gave their views freely and uninhibitedly despite constant direct and indirect reminders to stay on the topic from me as the moderator. The excerpt from one of the focus groups (TURNS 126-133 FG One, Appendix 4.3) illustrates this.

Evaluative criticism of the materials came through at each phase, which was detailed and insightful. When the session drew to a close I thanked the participants and gave a debriefing about the nature of the research. Ethical

issues of research were shared with them again. Participants were encouraged to keep confidential what they heard during the meeting and I assured them of the fact that I would be anonymising data. After the meeting a summary of the prevailing mood and critical comments about the session, including representative quotes, I wrote down in my research diary.

3.7.3.2 Procedure for analysis

Literature on focus groups suggests two approaches to content analysis, the first being the relatively simple and subjective *cut and paste* technique and the other more rigorous use of a variety of specific methods and techniques which emphasise reliability and replicability of observations and the resultant interpretations, (Stewart and Shamdasani 1990). The *cut and paste* technique is another name for what Krueger and Casey (2001) refer to as the *long table* approach. In this method transcribed data is thematically colour coded and then all the related or same coloured strips are cut and grouped together under one category. The same can be done by the cut and paste function of a PC.

I analysed the focus group data at different levels using different approaches. Verbatim complete transcription was undertaken and the points respondents made were also lifted from the raw data and recorded. In effect data was prepared using both the *cut and paste* technique and the content analysis approach where data is reduced to categories and keywords.

The literature on focus group analysis suggests that when conducting analysis, consideration should be given to the following five factors: words; context; internal consistency; specificity of responses; emergent themes (Merton *et al.* 1990). At word level, actual words and meanings need to be determined by a frequency count of commonly used words, with the idea of clustering similar concepts together. Such an analysis can be conducted with a concordancer or NVivo. However, this was not a major focus in this research because I perceived it more fruitful to focus on opinion expressed than on words used. But when certain words and expressions impinged on the nature of the interaction it was noted. Moderator prompts provided a natural context for the use of certain expressions and words which acted as the triggering stimulus for the discussion. But from within the utterances of the group certain expressions acted as a stimulus for further discussion, for instance expressions like “subconscious learning” and “scoring scheme” assessment (Appendix 4.3: FG One).

Internal consistency was maintained by tracing the flow of conversation including changes within that particular thread of discussion and reverses in positions that participants took in response to what someone else may have said. (For example in the exchange in TURNS no: 108-114 (Appendix 4.3) the utterance in the middle about “sound effects” introduced the word ‘boring’ which was picked up by others.)

Specificity of responses was taken into account by giving more weight to responses that were specific and gave concrete examples from personal experience of working on the materials. Responses that were vague and impersonal were given less weight (Appendix 4.3 TURN No: 288 FG one).

Finally points that were made repeatedly which were specific to the CD-ROMs were categorised. The main trends and overarching themes spreading across the entire discussion were identified and the data coded and categorized accordingly.

3.7.3.3 The findings

The findings from this method are reported in Chapter Four: Focus Groups as an Evaluation Method. Anonymisation of reporting was done and the participants of Focus Group One (FG1) were anonymised as students A,B,C,D,E,F (f)and G, and Focus Group Two (FG2) as H,J,K,L,M(f),N(f),O(f) with (f) denoting female gender.

3.7.4 Retrospective protocols

The version of think-aloud protocols/verbal protocols used in this study is retrospective in the sense that students finish one section and make a recording of their impressions. These intermittent recordings are not strictly post-event as retrospective protocols are meant to be. They are an ongoing contiguous hybrid of think-aloud protocols and retrospective protocols. A fuller description of introspective and retrospective methods was given in Chapter Two Section 2.6.2. Here, the lessons from the pilot study which impinged on the hybridization of the method for the main study are discussed first, and then I shall describe the use of the method in the main study.

3.7.4.1 Emergent design

The design which was used here evolved from a marriage between a ‘paired user walkthrough’ and a ‘think aloud’ session in the pilot study. The experience there, where both these methods were tried out, suggested that

the participants found it difficult to verbalise their thoughts as they worked with the materials, even though they were given a demonstration of how this could be done, and they were asked to practise thinking aloud before recording their thoughts. The reason could be attributed to their involvement with the materials and insufficient practice with the method. The method was repeatedly explained to them but the spontaneous concurrent nature of TAP was not captured. What emerged was a concurrent retrospective account, recorded periodically, but only after repeated reminders. This reluctance to record in the conveniently provided tape recorders also suggested that participants were not naturally eager or inclined and had to be somehow motivated to record their thoughts.

The experience of the paired user walkthrough in the pilot study also appeared to be similar. The students were encouraged to talk to each other about their experience of working with the materials. They were told that any time that they found things difficult they should verbalise their thought processes, and that when they found an exercise particularly interesting or boring they could comment on it. Again they had to be constantly reminded to talk and record. According to Ericsson and Simon (1987) this can be a necessary condition even after subjects have received extensive training. Students' involvement with EASE and also their desire to work on the materials undisturbed resulted in the pairs splitting up and eventually working at individual work stations. When prompted to talk together they did so with reluctance. There was no option but to allow them to drift apart and only one pair stayed together.

Where students came from the same linguistic background they were told that they could record their thoughts in their mother tongue if they felt inhibited using English. One pair did talk in Chinese and two students occasionally in Japanese. Most of them however felt that they should talk in English as it would give them much needed speaking practice. In that sense, there was conflict of interest at work as well which impeded the TAP or the paired walkthrough to proceed in a textbook-perfect fashion. The students were more interested in their own development rather than fulfilling the objectives of this research. This is to be expected and had to be accounted for in the design of the main study.

In the main study a revised version termed retrospective oral protocol was used.

3.7.4.2 *Retrospective oral protocols*

The Retrospective Oral Protocol (ROP) procedure for the main study was designed so that all students could work independently, recording their thoughts as and when they found convenient. The paired user walkthrough was abandoned altogether after the experience of the pilot study and the concurrent nature of TAP evolved into a series of retrospective verbal accounts of what the student did between the first and second recording sessions. It was observed that students recorded their views more frequently in the early sessions (WAV Sessions 1 and 2) than in the later sessions (Sessions 10-12). As the course progressed the number of recordings became fewer either because the students became familiar with the software and had fewer significant events to report or, they thought it would be repetitive or, as in the pilot study, they became too involved in their own learning.

Student-participants were asked to think about the following questions:

What did you find interesting about this program and why?

What were your feelings / thoughts about doing this exercise?

Did you find it difficult or easy? What was difficult? What did you like about it?

Since in the training sessions I had emphasized the evaluative aspects or had explained the *purpose* for this recording the students used evaluative language and expressed their stance in relation to the materials. They also reflected on their own performance but these comments often degenerated into accounts of what they did or had been doing, without any evaluative element. The students did express their opinion of the exercise and whether they found it easy or difficult, and this served my purpose well.

Unlike the verbal protocols used to identify learner strategies in Legenhausen and Wolf's (1990) study, here it was felt that the students should stop to record only when they had something of significance to report. This approach is corroborated by HCI literature where in user walkthroughs or cognitive walkthroughs users tend to report only when there is something significant, like an error for example (Preece *et al.* 1994).

WAV is the format for storing sound files in Windows and is the de facto standard for sound on PCs. WAV sound files have a wav extension appended at the end of the file, and can be played by nearly all windows applications that support sound. WAV was selected because of its ubiquitous availability and system

compatibility with PCs in the laboratory. Most of my students were also PC/ Windows users rather than Apple /Macintosh users so their familiarity with the system and ease of use were also considered.

The students were provided with microphones or they used the built-in microphones to record their impressions and uploaded the sound file to a designated folder on the designated space for the course on one of the university's servers. These folders were labelled with their names and student identity numbers. They were copied onto CD-ROMs and later transcribed and copied into QSR NVivo Version 2.0.161 (the qualitative research data analysis software).

There were around 50-53 students in the first two sessions, the number settled to 45 in the third week. Therefore there are some comments coded by students who later dropped out.

Once in NVivo, a quick perusal revealed that the coding scheme used to analyse the focus group data could be applied initially to the data from the protocols. As coding progressed more themes peculiar to the ROPs were revealed, and these were added to the original list.

3.7.4.3 Retrospective reflective written accounts

Retrospective Reflective Written Accounts (RRWA) were not used in the pilot study. It was noted, however, that the subjects of the pilot study might have given more measured and thoughtful commentaries on their work if they had not been working under time pressure. The subjects had to finish the exercises in the unit that they were working on, and they wanted to get on with what for them was the real business of learning. Reflection is considered an integral part of experiential learning and is instrumental in personalizing learning (Boud *et al.* 1985). It is also perceived as encouraging independence and autonomy in learning (Kolb 1984:68-69). Through reflective practice the students take charge of their learning, thus strengthening and fortifying their experience (Watkins *et al.* 2000:93).

Since the EASE materials were being used as part of a course on communication skills, there were certain course requirements that had to be fulfilled. Written and spoken output was necessary in order to assess the students' learning. I asked the participants to write a retrospective reflective account after each of their 2.5 hour laboratory sessions and send it to me by email. Each student sent between 8-10 emails which were filtered to a designated folder in my mailbox. These emails served the purposes of both this research and the course

requirements by being counted towards the students' class participation grade. As mentioned earlier, because LUMS students are high achievers and thus highly 'grade conscious' this strategy worked well as a means of maintaining motivation.

The procedure for the analysis of these e-mails involved entering them into NVivo under the students' names and the date on which it was sent. Then this data was coded according to previously established codes which had already been applied to the oral protocols. Any new codes that emerged were added.

3.7.4.4 The findings

The findings from this method are reported in Chapter Five: Retrospective Protocols. Anonymisation of reporting was done and the students were given numbered codes with the initial of the chapter so first student on the list was referred to as Student RP 1(M) through to Student RP 48. The alphabet (M) denoted males and (F) denoted females.

3.7.5 PLUM and SUMI questionnaires

A general description of questionnaires is given in section 2.7.3 of chapter two. The questionnaires PLUM (Programme on Learner Use of Media) and SUMI (Software Usability Measurement Inventory) used for this study were not designed specifically for this study nor were they adapted. Instead copyright versions of the questionnaires were used so that the validity and reliability of the instruments was not in question. The two questionnaires, PLUM and SUMI, were taken from credible resources, with established credentials. However the perspective of these two questionnaires is significantly different. The PLUM questionnaire was taken from the Open University's IET Evaluation Resource and is meant for use by teachers working with educational technology and needing to evaluate particular software for use in their practical teaching lives. This brings PLUM close to the context and requirements of this study. SUMI on the other hand is closer to the domain of computer professionals and is based on HCI theory and practice. It is also a highly standardized usability measuring instrument which is commercially available for software development firms to test their product for certification purposes (for example for the ISO 9000 certification of quality). One basic difference between the two questionnaires is that PLUM focuses on the learning effectiveness of the software content and the learner's experience of it, whereas SUMI is used to measure usability aspects of the software and is more generic in measuring the satisfaction and effectiveness of the program being evaluated.

The reason I used two questionnaires was that I wanted to take account of differing contexts that teacher-evaluators may be working in and the importance and scale of the purchase that an evaluator may have to support and validate through the evaluation findings. In this part of my research I also wanted to explore and represent more than one conveniently available questionnaire type. Most significantly, using two questionnaires enabled me to address concerns about the validity and reliability of this particular evaluation method. It would not have been possible for me to fully address these concerns had only one kind of questionnaire been used. The two selected questionnaire types are commonly used in educational settings. There exist, however, many other kinds of questionnaire, and the two selected types cannot be taken to represent all the different questionnaire types that abound.

3.7.5.1 PLUM questionnaires

The two pre-program use and post-program use PLUM questionnaires can be found in Appendix 6.1 and Appendix 6.2. The pre-program use questionnaire was administered in the first session of the course and the post use in the last session. The number of respondents on all occasions was 40-45 but a few incomplete questionnaires had to be discarded so the number of questionnaires for both CD-ROMs was 36-40 (a range is given here as each questionnaire had different numbers of respondents). The PLUM pre-program use questionnaire has only one open ended question.

3.7.5.1.1 Post use questionnaire

The post use questionnaire is a mixed design questionnaire which uses different questioning styles. There are four questions, but these have subsets. Questions 1, 3 and 4 are open ended questions and lend themselves to qualitative analysis, while question 2 has a battery of seven sub questions which require answers on a Likert scale. The first question refers back to the pre use question. It has three closely linked parts which encourage respondents to answer specifically. Question 3 again has four sub questions whereas the final question is a single *why or why not* open ended question (Cf: Appendix 6.2). Since both of the EASE CD-ROMs were being evaluated students were given two sets of the same questionnaire in succession. The analysis of open ended questions was conducted using NVivo; codes based on the points students made are counted and presented in Chapter Six.

Question 2 of this questionnaire was analysed using Microsoft Excel (2003) Pivot Chart/Table functions were used to reduce and analyse data giving tables and bar graphs of answers to each sub question. There were only seven variables in Question 2 presented in Chapter Six section 6.2.3.2 (for EASE *Listening to Lectures*) and section 6.2.4.2 (for EASE *Seminar Skills*). Furthermore, the *Cronbach's Alpha* reliability tests were conducted on the data from PLUM to determine how internally consistent respondents were in answering the questions and to determine whether the respondents understood the questions in exactly the same way as they were meant to be understood. This was done to show that data from the PLUM questionnaire was not taken at face value but statistical measures were applied to confirm the data's validity and reliability.

Reliability test of the measuring instrument PLUM post use questionnaire Q2

Reliability tests were also carried out to examine the internal consistency of the measures (scales) used in the questionnaire². This was undertaken by using Cronbach's alpha test. Churchill (1991:68) states: "coefficient alpha absolutely should be the first measure one calculates to assess the quality of the instrument". Alpha values of 0.7 and above are normally regarded as acceptable to good (Hair *et al.* 2005). In this study the alpha values 0.79 for EASE *Listening to Lectures* and 0.89 for EASE *Seminar Skills* thus confirming data reliability (Cf: Appendix 6.3/6.4)

3.7.5.2 SUMI

The SUMI questionnaire (Kirakowski and Corbett, 1994) has standing in the HCI industry. This description of it is based on the handbook that accompanies the inventory (Kirakowski 1998). The actual implementation and discussion of how the SUMI results are meant to be interpreted are given in Chapter Six, alongside the actual results, to make analytical discussion and cross referencing more straightforward.

² V1: Easy to operate
V2: Enjoyable to use
V3: Provides good support for the exercise
V4: Provides good advice on how to approach the task
V5: Helps you learn
V6: Fits well with the rest of the course
V7: Well worth the time spent on it

Kirakowski (1998) states that SUMI can help evaluators gather objective data about software after end users have evaluated the program. SUMI has been used to measure satisfaction of a variety of programs in different situations including space station control as an extreme example. Respondents normally take ten minutes to complete the inventory (Kirakowski 1998). If the users have no prior experience of the software, however, more time may be required for introduction and training. Data from a SUMI evaluation is compared to standardised expected results to establish its usability and user satisfaction.

Since this study's methods are evaluating CALL MM materials, the use of HCI domain specific evaluation criteria is justified. According to the definition of usability given in ISO/DIS 9241-11(1995) 'Guidance on Usability' a method can be evaluated on the type of measures that it provides (Dix *et al.* 2004; Preece *et al.* 1994; Kirakowski 1998). These measures can be 'Effectiveness', 'Efficiency' and 'Satisfaction' and can be analysed quantitatively or qualitatively depending on the nature of the data they yield (Dix *et al.* 2004). These usability measures mentioned in ISO/DIS 9241-11 (1995) are taken as yardsticks to evaluate whether methods of this study measured usability by measuring Effectiveness, Efficiency and Satisfaction. (Cf: section 3.8.7 for more detail).

3.7.5.2.1 Aspects of user satisfaction

In discussion of the key attributes that usability should measure Porteous, Kirakowski and Corbett, (1993) identified five aspects, which they called aspects of user satisfaction. Kirakowski (1998) sees close correspondence between these five scales to the seven dialogue principles of ISO/DIS 9241-10, with four being identical. They take 'satisfaction' as the overarching attribute which can be divided into further five aspects. They use these aspects in creating SUMI scales. The following definitions of these five attributes of user satisfaction are after Preece *et al.* (1994:520-1) and Kirakowski (1998:5)

Efficiency: this refers to the feeling of the users that the software is enabling the task(s) to be carried out in a fast, effective and economical manner or, at the opposite extreme that the software is getting in the way of performance (Kirakowski 1998:5).

Affect: this is a psychological term used to describe the emotions attached to an action. In the context of usability evaluation it refers to the user feeling mentally stimulated and enjoying a pleasant experience, or the opposite as a result of interacting with the software (Kirakowski 1998:5).

Helpfulness: this refers to the perceptions of the users that the software interacts and communicates in a helpful way and assists in the resolution of operational problems if they occur (Kirakowski 1998:5).

Learnability: this sub-scale refers to the feelings of the user that the software is relatively straightforward, and that it is easy to become familiar with the software and that its tutorial interface, handbooks etc. are readable and instructive (Kirakowski 1998:6).

Control: this sub-scale refers to the user's feeling that the software responds in a normal and consistent way to input and commands, and that the software is not difficult to operate, enabling the user to get their work done easily (Kirakowski 1998:5).

The five SUMI scales measure these aspects. (There is also a sixth scale, called 'Global usability' which is a general satisfaction measure, Kirakowski 1998:6). After SUMI has been administered and the data run through the special statistical analysis program (SUMISCO) these scales provide 'Profile' of the software. Kirakowski explains that these SUMI scales make up the user satisfaction 'profile' of a software system (1998:6). "Unlike in other approaches, the SUMI scales have been discovered through analysing thousands of records of users' data using 'factor analysis'(Kirakowski 1998:6). He further asserts that scales have been "checked and replicated numerous times" thus reassuring users of validity and reliability concerns and the robustness of the inventory. He explains that that these scales purport to measure the "...the degree of internal balance between the demands of the task and/ or the computer system and the level of knowledge or the ability of the user." (Kirakowski 1998:6).

3.7.5.2.2 Validity and reliability of SUMI

The validity and reliability of the SUMI scales has been established in repeated studies. One such experimental study reported by Molich *et al.* (1998) upholds SUMI results when compared with inspection methods. In this study four teams evaluated the same software, two teams using inspection methods, one team using SUMI and the fourth team using a mixture of inspection methods and SUMI. The SUMI results were remarkably similar

and consistent, thus establishing its reliability, and the *Item Consensual Analysis* function of SUMI identified exactly the same problems that the inspection methods teams had agreed on, thus establishing SUMI's validity.

SUMI was selected for use in this study because it is a time tested usability tool. Preece *et al.* (1984) refer to SUMI as "an internationally standardized questionnaire" and mention the above cited five subscales on which the SUMI question batteries are based. For the purposes of this study it was felt that there was no point in reinventing the wheel by designing a new questionnaire from scratch. Validity and reliability, as stated earlier, have already been established by initial extensive prototyping and frequent reiterations. (Cf. History of SUMI Development: Kirakowski 1998: 59-70).

3.7.5.3 The findings

The findings from this method are reported in Chapter Six: Questionnaires as Evaluation Instruments. Anonymisation of reporting was done and the students were given three digit numbered codes preceded by the initial of the chapter. So each student on the course was given codes starting from Q-001 through to Q-040. The three digit format was used to accommodate SUMI's analysis requirements.

3.7.6 Activity monitoring and key stroke logging

The method of electronic observation and studying tracking data captured through activity monitoring was described in section 2.6.4 of Chapter Two. The rationale for using this method in this study is to determine how exactly learners work with this program and indirectly to support the validity of findings of other methods. Data from activity monitoring can be objectively collected and analysed with little intrusiveness. An objective method of observing student behaviour and interaction with the materials provides evidence which either supports or erodes claims derived from data gathered by other methods. It will determine whether there is a match between what subjects say in the focus groups or RP or questionnaires, and their actual practice.

Monitoring software, also known as spy software is routinely used by organizations to keep track of their employees' online and computer based activity. The kind of software that was used for this study was the type that monitors and records information about individuals who could be employees in an organization or children working on computers at home. It is software that can be installed remotely through the LAN on all PCs connected to the network. Two web based companies activity monitoring products were used in this study. A

trial version for the pilot study was downloaded from ACTMON (<http://www.actmon.com/>) and for the main study the multi-user licensed software *Computer Monitoring Internet spyware and Keylogger* was purchased from *Softactivity.com* (<http://www.softactivity.com/>). The *softactivity.com* has the following uses listed on its webpage

- *Employee activity monitoring. Find out what they are doing when they are assumed to be working*
- *Control over students during the academic hours. Easily track their activity, so they will use network for the educational purposes only*
- *Parental control. Our powerful spy software will alert you when your child is into something suspicious.*

Screenshots of the students working on the materials were recorded at 15 seconds intervals by the activity monitoring software. The time bar at the top end of the computer screens showed actual time and was captured in the screenshots. This pictorial medium of the screenshots was a much richer medium of observation than the textual reports of the key logger. Key logger reports were sent periodically by the activity monitoring software to the designated space on the LAN server.

In the pilot study conducted at Warwick, a trial version of the activity monitoring software was used. This 60 days trial period from ACTMON was considered adequate to collect the data. Due to some conflicts with other programs the monitoring software would not always work, however, and only four of the eleven terminals used for the study sent regular periodic activity reports to my email address. Because this was a pilot study the sample of collected activity reports was considered adequate. For the main study these software conflicts would be resolved.

For the main study the activity monitoring software was set up on 45 terminals and programmed to retrieve data for the times of the class (1730 to 1930 hours). All students were informed that their activity on the computer for the duration of the class was being monitored and recorded using computer activity monitoring and key logger software.

Analysis procedure for key logger and activity monitoring was complicated as far as interpretation and representational format was concerned. Flowcharts using VISIO professional flowchart making software were made of the two EASE CD-ROMs. However as this was a time consuming activity. Unit 2 and Unit 6 of

Listening to Lectures and Unit 1 and Unit 5 of *Seminar Skills* were selected for purposes of making flowcharts to determine the navigation patterns and time frames students worked in. The data from 40 students could not possibly be thus analysed so 20 students' record of screenshots and key logs from the activity monitor program were used to make flowcharts of the two units they worked on the CD-ROMs. This selection was random. According to Hofland and Johansson random sampling enables the selection of individual texts "free of the conscious or unconscious influence of personal taste or preference." (1982: 3). On the other hand, Richards (2003) suggests that the nature of qualitative case studies lend themselves more to purposive sampling. As I was interested in exploring all kinds of students' work, not just the good ones, I decided to make a random selection using Microsoft Excel. Cases were selected on the basis of their "*instrumental*" focus as opposed to their "*intrinsic*" worth (Richards 2003:21).

3.7.6.1 The findings

The findings from this method are reported in Chapter Seven: Tracking and Activity Monitoring as an Evaluation Method. Anonymisation of reporting was done and the students were referred to in this chapter by the number of the PC they were working on. The original computer laboratory PC allocation list was used to anonymise the students' names. In Chapter Seven PC53 is a female student who is shown working on unit two of the first CD-ROM and similarly all reference to selected students' work is referred by their PC names. Each data study chapter has used a different anonymisation code to differentiate between the data of each chapter.

3.8 Criteria for evaluating methods for multimedia CALL materials

During the course of planning, piloting and engaging in this study, I came to perceive that in order to gauge the efficacy of the methods of evaluation a set of criteria was needed on which the methods could be assessed. My purpose was to identify methods which focussed not just on the functionality of the MM materials but also captured their impact on learners and the learners' responses. The chosen criteria needed to reflect this purpose. Moreover more practical considerations such as the feasibility of use of methods by potential selectors of materials also had to be kept in mind. These eight criteria are a composite set of what has been identified as key attributes of usability evaluation methods in HCI and educational research methods. All these attributes are also potentially applicable to evaluation methods for paper-based materials except for the 'usability measure' which is applicable only to electronic materials. The methods used in this study were evaluated on the following criteria

but these criteria can be used by teacher evaluators-selectors for assessing other methods as well. These then are the criteria.

3.7.7 *Cost effectiveness*

Time/Resources/Finance

One consideration in selecting any evaluation method is the amount of time and money required to conduct an evaluation. Time can be translated into monetary terms if we consider the teacher/evaluator devoting many working hours to carry out the evaluation. So if the method requires extensive time lay out then it may not be cost-effective and the teacher evaluator may have to justify the time spent with the special reasons (what special value is being added) that merits the use of such a method.

A method can be expensive or inexpensive in terms of both time and money depending on the level of technology. Audio and video recorded observation along with computer logs or tracking techniques can be expensive in terms of both time and money (through cost of equipment and transcription) whereas if paper and pen are used to take observation notes, or to complete user notebooks instead of computer logs, it can be relatively inexpensive. Dix *et al.* (1998) make this distinction when discussing the appropriate choice of usability evaluation methods.

3.7.8 *Ease of use*

Pre use preparation and implementation and post use data preparation

Preparing to conduct an evaluation will entail time and money considerations but will also have a cognitive dimension. A method may involve extensive training of the evaluator or of the respondents or may require considerable time setting up a laboratory or training laboratory assistants. The learning curve for conducting an evaluation using a particular method may be steep or shallow for both the evaluator and the respondents.

In the implementation stage of the evaluation an important consideration is the ease with which a method can be implemented. Methods which employ state of the art software or systems require greater expertise and level of familiarity with the software and the system used.

Preparing data for analysis in the post use phase of some methods may prove to be a difficult exercise and the teacher evaluator may have to undertake training on how to do this. Transcription of interview data may be time consuming but is still relatively easy when compared to learning how to use special statistical analysis tools for questionnaire data. Thus preparation time varies according to the method.

3.7.9 Bias: Researcher or Respondent

Another important consideration for an evaluation method is the level of bias that may creep into the data. Cohen *et al* (2003), Nunan (1992) and Dix *et al* (2004) refer to this as the level of subjectivity or objectivity of a method. For instance data from controlled experiments have a high level of objectivity as compared to data from more subjective techniques such as think-aloud retrospective protocols. Data from subjective methods are likely to be more vulnerable to bias because researchers can influence the results by their own preferences. However, even if the raw data is more or less objective the interpretation of data can exhibit a certain level of subjectivity because it is being filtered through the prism of the researcher's orientation or preferences.

3.7.10 Ecological validity / Intrusiveness or Interference

Ecological validity (Preece *et al* 1994:698) is the degree to which an evaluation will affect the results obtained from it. The interference of the situation and the control exerted by the evaluator on a user's tasks can affect the data obtained from the evaluation (Preece *et al* 1994). When a method allows or admits the behaviour of respondents to be affected by the presence of an evaluator or observer this is referred to as "intrusiveness" (Dix *et al* 2004). Intrusiveness can also occur when the observation is being recorded by a machine like a computer or video/audio recording device. Some users' behaviour may appear contrived and unnatural because of this intrusion and this may affect the results of the evaluation. Preece *et al* (1994) suggest that the artificiality of the situation created by the method should be taken into account when analysing data.

3.7.11 Nature of information

Deep-Surface/ Intrinsic-Extrinsic

Information yielded by a method can be at differing levels of depth or conceptual density. Dix *et al* (2004: 359) illustrate this principle with an example: they see information as low-level when it comments on the colour schemes and font styles or effectiveness of iconic representation used in the graphic user interface (GUI), whereas they regard high-level feedback as impressionistic or analytical commentary on the effectiveness of the

system. Preece *et al* (1994:698) on the other hand refer to this criterion as the “scope of information needed” and they define it as “...completeness of the data collected in relation to the eventual use of the system”. For the purposes of this study it may be more appropriate to categorise the nature of information gathered as ‘low surface level’ or ‘deep high level’. These categories correspond to Bloom’s (1956) taxonomy of question types where the classification is in terms of complexity, from the low ‘recall’ level of knowledge to the advanced level of evaluation. Surface (extrinsic) or low level information could be information regarding aspects of GUI design and the general impression of the program’s layout (in Dix *et al*’s terminology, 2004:359). Deep (intrinsic) or high level information might include comments on, for example, the correlation between the pedagogical objectives of the materials and the content or style of exercises employed. These might be specific comments or an overall impression of the effectiveness of the materials. Hémard (2004) gives a more interpretivist slant to the issue of “nature of information” and refers to it as ‘quality’ of data which he interprets as the qualitative or quantitative nature of the data revealed by a certain method. Hémard’s (2004) terminology appears to be more subjective and interpretive as compared to the more positivist approach taken by Dix *et al* (2004).

This criterion is particularly important for my target evaluators because it will contribute particularly to the possibility of determining the appropriateness of materials for their learners in their learning context.

3.7.12 *Immediacy of response*

This criterion deals with how quickly a method yields discernible feedback after use i.e. whether results are immediately available from an evaluation method or have to be reviewed and prepared extensively before some sense can be made of them. Dix *et al* (2004:359) identify this as having implications for the efficiency of a particular method. Observation methods which use activity logs and video recordings can delay the results because the data can only be obtained after the users have stopped using the PC which was monitoring their activity or tracking their use. As analysis can take considerable time, the results would not be immediately available from such a method. Low technology observation, using perhaps just paper and pen, yields data right away. Similarly focus group data is immediately available. This particular criterion is important because of the flexibility it offers to the evaluator, methods which rate as having greater immediacy of response can be used if time is short or a quick, general overall impression is required. An evaluation report can be prepared later, from the same data, if necessary.

3.7.13 Usability measure

Satisfaction/Effectiveness/Efficiency

This consideration relates to the effectiveness of a method in measuring important aspects of software design, functionality and instructional aims (Satisfaction, Efficiency and Effectiveness). This criterion is peculiar to computer based materials whereas the other criteria can also be applied in some way to the evaluation of paper based materials. The term 'usability' is used here with the meaning it is assigned in HCI literature, and is not to be confused with 'usability' in its more general sense, which can be equated with 'user-friendliness' and 'ease of use'. In this list of criteria "Ease of Use" is a separate criterion which evaluates the method and not the software and is distinct from software usability. (A more extensive discussion of usability was undertaken in an earlier section 2.4 of chapter two.)

The sub-criterion of *Satisfaction* is considered important from the user's perspective (hence its importance to this learner/user centred study). It purports to measure how users get along with the software. Satisfaction is an internal state variable and aims at measuring the strength of feeling of an end user. Satisfaction would be low if the users experienced stress or irritation, or if the content did not match their ability. Low satisfaction would be registered as lack of concentration with a tendency to make elementary errors, and constant assistance seeking (Kirakowski 1998:4).

The sub criterion of *Efficiency* is a user performance variable which aims at measuring the speed and time with which the user completes a task. Efficiency is considered the easiest aspect of usability to measure (Kirakowski 1998:4). Smooth snag free functioning of the software's functional features could also be part of this measure.

The sub criterion of *Effectiveness* aims to measure output in terms of appropriate and correct use of the software by the user as it was designed to be used. The user's use of the materials, such as following instructions the way they were meant to be followed, is compared to the predetermined desired output features of the materials. This can involve quite detailed analysis and can become subjective as each user may navigate or work the software/system in a different way, according to personal preference. This criterion can be very important for my target evaluators, however, as it will relate to how materials are actually used, whatever the creator's intentions.

3.7.14 Robustness of method

Validity/Reliability / Productivity

The rigour and robustness of a research method is established by the validity and reliability of the data that it elicits. The same is true of evaluation methods.

According to Nunan (1992), Preece *et al* (1994) and Seliger and Shohamy (1989) the validity of an evaluation method means whether it measures what it purports to measure with respect to the specific purpose of the evaluation. Preece *et al* (1994: 697) suggest that this could mean two things. Firstly, validity can indicate whether a method is valid for a given purpose (for example if users' attitudes are to be determined then the users rather than the experts should be the evaluators). Secondly, validity can indicate whether the measurement is valid (for example only experts can get the right measurements using certain methods and these methods should not be used with novice operators).

Consistency or replicability is essential for reliability (Nunan 1992; Preece *et al* 1994; Cohen *et al* 2003). A reliable evaluation method produces the same results on separate occasions under similar circumstances. Carefully controlled and executed experiments yield reliable data particularly when carefully selected subjects perform tasks and the interaction is carefully controlled and monitored (Preece *et al* 1994:697). The degree to which the method elicits accurate data will govern its reliability (Seliger and Shohamy 1989: 184).

The amount of usability related issues and other significant points about the materials that a method generates, or identifies, or assists learners in identifying, can be considered as the method's productivity. However this will only be effective as a comparative scale and as with the other methods, the findings will be dependent on the skill and expertise of the analyst (Lavery, Cockton and Atkinson, 1997).

This criterion may be the most useful in determining the worth of a method for my target evaluators because the more productive the method the more issues it will identify. My target evaluators may well be non-experts without the requisite specialist skills to interpret the findings, but they are likely to want to identify the maximum number of issues for the purposes of summative evaluation. Interpretation of these issues is more likely to be the mandate of the expert developer conducting a formative evaluation.

3.8 Summary

This chapter began with the research questions of this study. The theoretical and conceptual framework and the emergent research design were then discussed. The reasons for selecting a number of methods for trialling in the pilot study were given, followed by an account of the pilot study, a discussion of its findings and a complete description of the main study. The final section of this chapter established the cornerstone of the methodology of this study: the criteria which will be the basis for evaluating the chosen methods of evaluation. A two-tier approach to analysis will be undertaken in the next four chapters which are the data study chapters. In each chapter the data will first be analysed from the perspective of an actual evaluation and then the performance of each method will be analysed on the basis of the criteria established here.

We now turn in the next four chapters to an analysis of the data from the selected methods.

CHAPTER FOUR: FOCUS GROUPS AS AN EVALUATION METHOD

4.1 Introduction

The use of focus groups as an evaluation method was discussed extensively in chapters two and three. Chapter Three section 3.7.3 described how the method was used in this study. This chapter gives an account of how data from focus groups was collected, analysed and codified using three different approaches. The final section discusses the merits and limitations of focus groups as an evaluation method based on the criteria of method effectiveness established in chapter three of this study.

4.2 Focus group findings and analysis

The specific demands of this study required the analysis and reporting to be done from two angles. Firstly from the perspective of the teacher evaluator who is using focus groups to determine the effectiveness of particular MM teaching material and its instructional and functional content, secondly from the perspective of focus groups as an evaluation method for MM materials.

4.2.1 *Initial focusing exercise*

To help students focus on the goals of the evaluation process and to be able to gather relevant information a set of issues were set out as discussion topics in a rank ordering Focussing Exercise(Appendix 4.1). The following statements were used for the focussing exercise. These statements were specific to the two EASE CD-ROMs used and I myself developed them after determining the pedagogical objectives of the materials (see also Section 3.7.3.1 for the rationale).

Table 4.1 Focussing exercise statements

A.	Materials that provide an opportunity to familiarise your ear to different accents and ways of speaking.
B.	Multicultural classes and faculty are the norm in most universities and a student should adapt to this. Materials that take this into account.
C.	Materials that give adequate practice in academic skills particularly the skill of listening and speaking.
D.	Materials that give adequate practice in grammar, reading and writing.
E.	Materials that develop cognitive abilities by setting problems and tasks for students to develop the skill of listening because most university teaching is lecture based.
F.	Materials that teach note taking, paraphrasing and summarising skills.
G.	Materials that involve the whole person by including stories and humorous anecdotes that lecturers would use in their lectures.
H.	Materials that provide immediate feedback on learner's errors and suggest additional materials through hot links embedded in the CD.

The students debated these statements enthusiastically and were duly prepared for the ensuing discussion because most of the relevant aspects of the EASE CD-ROMs were brought to the fore through this discussion.

Table: 4.2 Results of the ranking exercise

Focus Group One Ranking Exercise Emergent Order: C E F H A G B D
Focus Group Two Ranking Exercise Emergent Order: E B C F A H D G

There is only a slight variation in the result of the two groups, CEF feature as the more important for both groups which are namely features of a CD-ROM or materials that develop academic listening and speaking skills followed by materials that develop cognitive abilities and note taking and summarising skills Choice B (Multicultural classes and faculty are the norm in most universities and a student should adapt to this. Materials that take this into account) showed up a difference in opinion between the two groups as focus group one put it towards the end and group two in the second position. Group two appeared not to have debated Choice B very thoroughly in the interaction.

This exercise not only achieved its purpose of easing the groups into debate but also threw up interesting evaluative comments on the CD-ROM itself. This warm up discussion provided a natural transition to the more specific and open ended questioning techniques that followed.

4.2.2 *Main focus group discussion*

The following set of questions was culled from the literature on evaluation and adapted to a learner centred approach to evaluation of MM materials. The broad topics around which questions were designed were: focus of the materials, pedagogical content and its organisation, instructional format, environment, presentation, functionality and online guidance. A general subjective and affective question at the end was about students' /participants' gain from the materials. This questioning scheme imposes a structure on the exchange and gives direction to what had to be discussed but does not in anyway inhibit the emergence of respondent led themes or constructs. The opinion expressed in answer to these questions is not controlled or influenced by these questions. Moreover, the language and structure of the questions was deliberately kept simple and reiterative to create a non-threatening and free-from-jargon environment for the discussion. Other related questions or clarification seeking questions, which the discussion prompted, were asked as the need arose.

Table 4.3 Discussion questions for focus groups

1.	Do you think the CD-ROMs fulfilled their aims and objectives, like did they have clear focus?
2.	What do you think of the structure and organisation of the CD-ROMs like is the material structured progressively in sequence?
3.	Do you think the exercises tasks are challenging? Are the instructions clear?
4.	What do you think of the instructional format is there a choice of levels?
5.	Does the learner have the opportunity for consolidation?
6.	Is simple language used?
7.	What do you think of the presentation of the layout of the CD-ROMs, is it eye catching colourful?
8.	What do you think of the working environment of the CD? Is it relaxed or distracting?
9.	Are they simple to operate e.g are user friendly?
10.	Are they easy to navigate?
11.	Is there any on line help on language available, like a dictionary?
12.	Did you gain anything from this experience? What did you gain?
13.	What are the best feature of the CDs?
14.	Did you learn anything that was unexpected

15. Would you recommend it to others?

These questions were designed to be asked in the order in which they appear but often the answers to some of the later questions came earlier as an off shoot to an earlier question or just emerged in the discussion. So in the interests of time I did not keep to the order of questions, particularly at the lower end of the list. This according to Morgan (1997) is acceptable because a “predetermined order of topics” need not be followed very rigidly. Merton *et al* (1990) are in favour of still greater flexibility in questioning and advise moderators not to adhere to using fixed questions and should skip over areas that have been covered earlier, because this may lead to discussion becoming more moderator dominated. By the same token questioning style and techniques can vary and be adapted from focus group to focus group of the same study. The pattern of questioning was adapted for the second focus group of this study.

To determine which aspects of the CD-ROM worked for this group of learners and which did not and the degree of consensus generated on each discussion point, data was analysed using three different strategies which operate at different levels of detail, description and depth of analysis.

The *first* strategy of data reduction was to reduce the discussion to the level of points or meaning units made by the student into broad categories and divided according to the basic criteria of where the participants agreed more and where opinion was divided. This was done based on an over all impression of a particular comment. The comments however were seen to fall in semantic groups. This strategy could be called *impressionistic analysis*. The *second* analysis strategy employed was to count the number of positive negative and neutral responses participants made to each question or prompt. This was called *pragmatic analysis*. The *third* strategy is the more common and detailed cut and paste technique with colour coding of main themes *thematic coded analysis*.

4.2.2.1 *Impressionistic analysis*

The main objective of this study is to evaluate the MM software from the point of view of the end users i.e. the students. In the focus groups the students presented their views on the software being used purely from their perspective although they do try to be objective about the merits and demerits of the materials on its own in their discussion. The participants were very aware that they were evaluating the EASE materials and used the language functions of evaluation to criticise and praise the materials.

The first analysis strategy was more impressionistic in its approach and the procedure I followed to get the information of value was to first reduce all the responses from both focus groups into meaning units or the points that students made. Then I made broad categories or groupings of these points to determine in which key areas the participants were in greater agreement and where opinion was divided.

Focus group participants offered a variety of opinions regarding the importance of various aspects of the program pertaining to the following categories emerged:

Table 4.4: Summary of emergent points

Matching the focus of the CD-ROMs to their needs.
Content specific criticisms and evaluations of exercises
Sequencing of learning content
Reality Checks: Rationalisation of negative criticism
Evaluation of the functional features of the CD-ROMs and coping with technology
Limitations of the program/ programming flaws
Environment of the CD-ROMs
Suggestions for improving the program
Managing the program in the context of the course
Motivational aspects of the CD-ROMs' interface
Comparisons with other CD-ROMs
Motivation through Testing
Motivation through feedback
Relevance to the needs of the students
Testing aspects
Grade oriented students
Recommending the program to others
Cultural Enrichment

However since the focus of this research is not so much on what comments students made on the particular CD-ROM being used but the nature of the evaluative comments and the frequency and saliency of certain issues raised over others, more emphasis was placed on interpreting the comments rather than on reporting them. This information was gathered impressionistically with greater emphasis on meaning rather than on counting. These points were gathered with close interaction with the data and despite being impressionistic are yet not subjective and would be of value to the teacher evaluator. But such a strategy would be difficult to practise for large volumes of data because very time intensive and the subjective nature of the analysis may raise issues of replicability. However, since no coding is done and data is allowed to speak for itself with direct interpretation of the students' responses the issue of subjectivity of coding categories when being done by a single analyser would not apply here thereby improving the analysis's validity and reliability. An impressionistic representation of areas of agreement and disagreement is given in Table 4.5.

Table 4.5 Areas of agreement and disagreement

Areas of strong agreement	Exercises and activities not challenging in <i>Listening to Lectures</i> .
	Note taking and summarising exercises best feature of <i>Listening to Lectures</i>
	Grammar exercises not needed
	Motivation levels low for <i>Listening to Lectures</i> high for <i>Seminar Skills</i>
	<i>Seminar Skills</i> CD-ROM very good
	Both CD-ROM user friendly easy to navigate
	Assessment and evaluation through a score board or marking scheme would improve motivation of learners
	Would recommend it to others
	Both CD-ROMs achieve the purpose of preparing students to study in Britain
Areas where opinion was divided	The working environment of the CD. Some members felt it could have used more colour and some sound effects.
	The sequence of exercises in <i>Listening to Lectures</i> .
	Vocabulary exercises very basic not in keeping with the level of intellectual complexity of the lecture content.
	The length of video clips.
	The effectiveness of <i>Listening to Lectures</i> in improving academic listening skills.
	Whether consciousness was raised as opposed to conscious learning having taken place

The above information gives the MM software evaluator the information s/he needs at a glance.

4.2.2.2 *Pragmatic analysis*

This method is more structured in its approach. It adapted the 'Long Table' approach to focus group analysis suggested by Krueger and Casey (2000). (Long table is a low tech simple method in which transcripts are spread on a long table or on the floor and with coloured markers and scissors themes are coded and cut out, can be done on the word processor and NVivo does it automatically). In my adaptation the first column had the TURN numbers the second column had raw data as uttered by the participant. The third had descriptive statement that expressed the meaning of the TURN. As a computer based analysis could also be done by identifying keywords in each utterance these were also identified. The procedure followed here was that the responses were first numbered and then reduced to points that the students were making and then whether this point had a negative orientation or a positive one was determined. There were some comments that were of a neutral nature. The grouping of responses was done according to the questions that were asked. So in answer to question number one there were N number of positive or negative responses in relation to the question asked. Similarly negative, positive and neutral comments were counted for each question. There is a danger here of researcher bias coming through because it is the researcher deciding whether a comment is positive or negative and a certain amount of subjectivity cannot be ruled out. After counting the assigned values, a clearer picture of which questions had more debate and where there was greater agreement or disagreement between members can also be ascertained. This is presented in the tables below.

Table 4.6: Biases in Learner Response

No:	Questions/Prompts	Statement Bias			
		Negative/positive/Neutral/Result			
1.	Do you think the CD-ROMs fulfilled their aims and objectives, did they have clear focus?	14	7	2	Negative
2.	What do you think of the structure and organisation of the CD-ROMs? Is the material structured progressively in sequence? Are the instructions clear?	4	12	1	Positive
3.	Do you think the exercises tasks are challenging?	6(CD 1)	9	0	Mixed neg Cd1 +Cd2
4.	What do you think of the instructional format? Is there a choice of levels?	3	5	0	Mixed
5.	What do you think of the working environment of the CD? Is it	2	6	0	Positive

	relaxed or distracting?		+5		
6.	What did you think of the content of the lectures?	1	11	0	Positive
7.	What about the technical features of the two CD-ROMs like is the operation simple and user friendly?	3	2	1	Mixed
8.	Does the learner have the opportunity for consolidation?	5	1	2	Negative
9.	Is simple language used?	0	7	0	Positive
10.	What do you think of the presentation of the layout of the CD, is it eye catching/ colourful?	1	5	1	Positive
11.	Are they easy to navigate?	2	7	1	Positive
12.	Is there any on line help on language available, like a dictionary? Did you use it? Any other links?	6	4	1	Mixed
13.	Overall have you learnt /gained from these two CDs?	3	14	0	Positive
14.	Would you recommend it to others?	0	7	0	Positive

Table 4.7: Negative/ Positive Results

Positive: 2, 5,6,9,10,11,13,14	Negative: 1,8	Mixed: 3,4,7,12
--------------------------------	---------------	-----------------

Based on the above analysis of the focus group interaction questions, the areas which showed greater levels of agreement between focus group participants can also be determined.

Table 4.8 Results

Areas of strong agreement	Areas where opinion was divided
Questions 1,2 5,6,8,9,10,11,14	Questions 3,4,7,12,13

This objective tabulation of opinion from the focus group shows that the overall results were positive despite the “halo effect” of the strong well worded opinion leaving a greater impression on the observer. This is not just a quantification of qualitative data for the sake of giving greater credence to the findings but I think the approach has real merit in an evaluation (Goulding 1998, Allan 2003).

4.2.2.3 *Thematic coded analysis*

The third strategy involved taking the long table approach suggested by Krueger and Casey (2000). This meant that the transcribed data was presented in the form of a long table reducing the participants responses to points or meaning units or descriptive statements and then further reducing these statements to categories. Then the data was revised again and these categories were applied to each response. The transcripts were tagged/coded according to these categories, (Appendix 4.3/4.4). These categories were further subsumed under themes. Each theme and sub theme or category was originally assigned a colour and the transcripts of the two focus groups were colour coded according to these emergent themes. Subsequent data analysis through NViVo eliminated the need for colour coding. At this point *purposive sampling* of responses was undertaken as opposed to the more positivist *random sampling* to select the most illustrative comments made by the participants to be included for interpretation in the discussion of each theme and sub theme/category. Moreover the entire quote or speech unit was selected and not words or phrases from it and at times the section of an entire interaction discussion point is given. The content analysis approach taken here is inductive where data reduction is moving from concrete to the abstract. Data reduction was conducted from the summary of points to categories to themes presented inversely in the following table:

Table 4.9: Themes and Categories

Themes	Categories	Definition
Learning Richness	Appropriatenes s	Relevance of the CD-ROM to the needs of students
	Content	Imbalance in level, and treatment or sequence of instructional content.
	Motivation	Self evaluation assessment mechanisms like scores to assess progress, and authentic content of video clips motivating
Improving Competence	Usability	Functional features of the CD-ROM navigation, online help, adaptability.
	Suggestions	Suggestions for improving the program
	Motivation	Features in design that affect motivation
Affective Dimensions	Management	Management of learning innovation
	Justification	Rationalisation of negative criticism
	Motivation	Motivational and affective dimensions of authentic cultural exposure

Three broad themes emerged from the focus groups transcripts. The first concerns the instructional content in the EASE CD-ROMs and its pedagogical design. The richness of the learning experience comes through irrespective of the negative or positive evaluative orientation of the comments. These important issues are

discussed under the broad heading of *Learning Richness* issues. The second main group of issues are grouped within the theme of *Improving Competence* and this theme includes subcategories dealing with functionality and adaptability of the user interface. The third broad theme has been identified as *Affective Dimensions* and comprises both the experience of learning through this innovation and reflection by the students of their own learning and their interpretation of what motivates them to learn through this medium. I faced some difficulty in making these categories or main themes mutually exclusive because in each one of these main categories *Motivation* as a theme kept showing up as a recurrent theme. So although there is a main theme named Affective Dimensions each theme has a sub theme of motivation related to that particular main theme. A detailed explanation and interpretation of these three themes and their subcategories follows.(Complete participant comments corresponding to the following section are given in Appendix 4.6).

4.2.2.3.1 *Learning richness*

Appropriateness:

One of the main concerns repeatedly expressed by the participants was the *EASE Listening to Lectures* CD-ROM's pedagogical objectives and their reception of its instructional content. They expressed concerns about the appropriateness to their needs of not so much the topics covered as the treatment of these topics. On the one hand they express appreciation of the content of the video clips and are full of accolades for the versatility and range of lecture content and the erudition of the lecturers; on the other they express dissatisfaction with the way this authentic material is exploited by the exercises that follow. Since this concern falls more under the ambit of the next subcategory *content* more detailed interpretation of this aspect will follow in the next section. The appropriateness concerns mainly stemmed from this misplaced sense of superiority to what was considered the target audience of the CD-ROM in their minds.

STUDENT D: it's like STUDENT G said English is almost like our first language....

This interpretation does not intend to dismiss the high level of proficiency of these students but to underline the attitude, which accompanies the teaching and learning of English in the context of this study. It is quite impossible for a researcher in an elite university in Pakistan to isolate English language teaching issues from the postcolonial dialectic of the prestige of English. Admitting to gaps in their knowledge of the English language would be to question the prestige and value of their private school English medium education. The gaps may be

evident to their instructors or even to the more proficient participants. However in the EFL/ ESL context of teaching and learning English these errors or shortcomings may not have been pointed out very rigorously to them. The more academically motivated students' comments show a certain self-reflection and a somewhat 'grudging' realisation of what they may have gained from the note taking and summarising exercises in *Listening to Lectures*. This reversal in opinion about whether they learnt anything from the exercise or not does question their concern of appropriateness of the CD-ROM *Listening to Lectures* content. The following comment illustrates this phenomenon.

STUDENT G: for our needs it depends on the audience. EASE would be brilliant for an audience which is not doesn't have the same perspectives... (Cf. Appendix 4.6 Table 4.6).

Another comment, which follows takes up this point:

STUDENT B: okay basically I think there are a lot of things that we take for granted ... (Cf Appendix 4.6 Table)

However, this criticism or concern voiced in the above quotes could also be a valid one because they are unanimous in their approval of the second CD-ROM *EASE Seminar Skills 1: Presentations*. The second CD-ROM they felt taught them more because it taught a skill to which they had little prior exposure. The authentic situations and the cultural enrichment aspect with greater focus on the student participants of seminars at Warwick was an affective factor.

The participants were particularly elated about the cultural exposure dimension of both the CD-ROMs and felt that this aspect particularly contributed to the richness of their learning experience. They felt that exposure to real language in an erudite context of a university was beneficial for them. One participant highlighted this when asked whether they had learnt anything unexpected from the CD-ROM by suggesting that "English English is quaint and funny and quirky" and EASE gave them an "insight into the quiriness".

STUDENT G: you know sprinkled all kinds of local flavour in it and English English is quaint and funny and quirky and everybody knows that and once you've mastered that idiom then you know you're qualified to go and study there and that's one of the advantages that it gives you an insight into the quiriness

This was an illuminating comment for it gave specific evidence of the richness of learning experience that these students, despite their concerns of appropriateness, felt they had gone through.

Since most of them aspire to study abroad either in the US or UK they felt that they had gained from that aspect of the CD-ROM. This aspect of richness of learning experience seemed to find currency with most participants of the focus group however enthusiasm may have varied from person to person. The following statement can also be taken as an example of intrinsic motivation and an appreciative response to cultural authenticity of these materials. Multiple themes can emerge from one quotation as is evident from the following excerpt. This excerpt underlines the themes of motivation and is also rationalising the accrued benefits and reflecting on the long term benefits of working on these materials.

STUDENT B: and I think one thing for sure that all of us hopefully after our Bachelor's at some point in time we will be hopefully going abroad...(Full Comment in Appendix 4.6 Table)

Content:

The instructional content of the materials generated most discussion, as that is what was contributing the most to learning richness. However a number of participants pointed out that they felt this sense of outrage in the first CD-ROM when the intellectual content of the lectures did not match the intellectual level of the exercises that followed.

STUDENT F (F): but after note-taking you are on this momentum that [?] so you're all pumped and you're like good good good...

This student goes on to talk about how there is this feeling of disappointment when they see the exercise after the lecture clip. The validity of this point is upheld when it is picked by the participants of the second focus group who also expressed their sense of being let down.

STUDENT O (F): what really kills you on CD-ROM one is that okay you're all tensed-up you know lecture lecture you sit listening carefully and like after two minutes of something world economic theory the question is how would you spell the world...

STUDENT N (F):...the exercises didn't relate to the lecture

Another concern expressed by participants was that the progression from easy to difficult although there, was very slow. Again they felt a certain amount of dissatisfaction at the first CD-ROM not being able to adjust to their level of progress, which only an intelligent CD-ROM would be capable of.

STUDENT L: [?] I insisted again and again that the level of the CD isn't really getting beyond a certain point it's pretty much staying at the same stagnant level except for I think the last unit.

This student later on qualifies his statement and reflects on his learning but still maintains that the level of progress was not what he expected.

STUDENT L: no but the thing is it teaches you exactly what content is necessary you know when you talk about listening to lectures ... (Full Quote in Appendix 4.6 Table)

Difference between the two CD-ROMs' content and teaching style was pointed out as well with *Listening to Lecture's* content having been already known and the novelty factor of the second CD-ROM *Seminar Skill's* contents making them more welcome.

STUDENT H: the first CD ... we didn't find it as challenging or we didn't learn as much as we could CD two ... (full quote in appendix)

The participants expressed concern over what they considered to be slow periods which they felt affected their performance on the CD-ROM and also had repercussions on motivation.

STUDENT F (F): but it becomes a bit monotonous

STUDENT E: And there were a lot of slow periods ...

Another aspect that made the exercises not so challenging was perhaps that the students felt that they lacked the academic rigour that they were expecting. The exercises failed to test their understanding but instead tested their memory at a superficial level.

STUDENT A: ...the lecture from which we had to take notes was so short that our note-taking skills was not actually tested ...

The issue of assessment and feedback on their progress kept recurring more in the first focus group than the second perhaps because the first group had more competitive grade oriented students.

STUDENT C: ...TOEFL that I took first it had an incentive to score high I mean I knew okay I'd score high on it but I wanted to score maximum on it I mean for example I was aiming at 300 out of 300...

STUDENT C: ...you scored 59 out of 60 you'd be like motivated okay like next section I'm scoring one hundred per cent

Again the comparison between the two CD-ROMs suggested that tasks and activities in CD-ROM-2 (*Seminar Skills*) were more challenging for this set of students.

STUDENT O (F): I think the exercise and the content of CD two were more interesting and they were better I mean you had to use your mind

STUDENT N (F): because they show you the classroom environment and people actually giving presentations

Moreover, the presence of Warwick students in a classroom environment attempting those same tasks gave a sense of relatedness and inclusion to these students in Pakistan.

Motivation:

The issue of motivation could not be separated from any of these themes. It is an ever-present yardstick that participants kept on applying to all aspects of the CD-ROM.

STUDENT L: I ...CD one so boring for us is the fact that we pretty much know what they are going to ask of us there is no level of surprise ...

The issue of motivation impinged on their impression of progression of difficulty levels in *Listening to Lectures*.

STUDENT H: I think like unit one and two were at one level then slight increase in units three and four and then slight increase in unit five and six and by the time you'd reached unit five and six you were so like tired and just bored out of your mind that you didn't want to do the [?]

However they also find the content of lectures in CD-ROM-1 interesting and recall the more humorous or intellectually stimulating lecturers with pleasure. These students appreciated the more subtle humour, which would have been lost on other students, for instance the Chinese students who worked on the CD-ROM in the pilot of this study.

STUDENT M (F): no some stories are funny that old man bio teacher...

STUDENT O (F): with the fish...

STUDENT M (F): the fish guy and the...

STUDENT N (F): the math guy's good...

They are evaluating some lecturers' ability to sustain their interest as opposed to others in the subject matter even though they may not have been students studying that subject. Or they may be studying it but the lecturer was not able to hold their interest.

STUDENT M (F): because the guy was interesting what he was talking about was interesting and he wasn't getting too technical ...

STUDENT M (F): no not necessarily there was this whole thing on AIDS that was really interesting but we're not studying AIDS I found it really interesting I was able to take notes

4.2.2.3.2 Improving competence:

This theme deals with those features of the focus group discussion that deal with the conceptual design of the application and its pedagogical objectives. This theme deals with both the technical features of the user interface, which mostly received a positive evaluation and suggestions of what these students considered ways to improve the effectiveness of the CD-ROMs.

Usability:

The competencies of the user interface were mostly positively appraised and the few 'glitches' in the functionality were pointed out.

STUDENT F (F): the logical structure is fine...

STUDENT N (F): yeah my scrapbook got lost a couple of times...

STUDENT H: and that's just a glitch in the program...

Suggestions:

The students gave some suggestions to improve the first CD-ROM and they related both to the sequencing of materials and the user interface. However there was mostly positive agreement in about the user interface and functionality of the CD-ROMs.

One student, whose comment follows, felt that there should have been more instruction. This student was one of the quietest ones in both focus groups and can be representative of the less proficient members of his class.

STUDENT K: ...they didn't describe the full exercises like I didn't understand the thesis questions ... (full quote in Appendix 4.6 table)

The following comment by the student is a suggestion about lightening the mood and creating more drama by including music or sound effects to the first CD-ROM.

STUDENT B: But I thought at times that if there were a little bit of sound effects with the exercises ...

Another student wanted more variation in the level and style of exercises and to break the monotony of same pattern in sequencing of exercises to bring in an element of surprise.

STUDENT F (F):... but it becomes a bit monotonous the units...

Another valid suggestion about sequencing and order of exercises follows which questions the entire pedagogical design of the CD-ROM.

STUDENT B: ...through the sequential order I think there should be note-taking should be moved a little bit up because basically we see that clip a zillion times before we actually come to the note-taking ...

Motivation:

The competence of the CD-ROM could be improved by placing greater emphasis on motivation by bringing in an element of surprise and by adding more feedback features of self assessment like scoring.

STUDENT F (F): ...was a bit more interchanging in the middle that would be good

In other words students would be more motivated if they were able to keep track of their progress and this would improve the over all competence of the CD-ROM. It would achieve more of its pedagogical objectives by improving the conceptual design features of the CD-ROM.

4.2.2.3.3 Affective dimensions

This theme deals with the affective dimensions related to how these e-materials were used in the context of the course these students were enrolled in and the assessment issues of this course. The issue that was debated often was about the need for an exam at the end, or an assessment in the form of an oral presentation, which tested the different components of the materials that they had been learning from. Other wider motivations or the long terms benefits like placement in universities abroad and familiarity with local language and cultural familiarisation are also categorised within this theme.

Management: Participants felt that they would have been more motivated had they had some means of evaluating their performance as an element of scoring and feedback in the first CD-ROM.

STUDENT C: give some scores back it was counted in unit one you scored 59 out of 60 you'd be like motivated okay like next section I'm scoring one hundred per cent...

The students felt they learnt more from the second CD-ROM on *Seminar Skills* because it taught skills which were core learning skills for them in relation to the course for which EASE was being used as supplementary materials. So the relevance to their need acted as a motivating factor.

STUDENT F (F): I mean he just likes the lectures and he wants to go to Warwick end of story no I think that the topics that were discussed were very interesting and the software itself is really good the first CD I didn't feel as in I found it good but I didn't feel that I learnt so much from it ... the second CD I thought I learned from more the presentation because everything we discussed before you know since we are at that point right now that we have to do presentations

These students also felt that since they were going to be assessed on their oral presentation skills they were motivated to learn from the second CD-ROM. This aspect ties in with the management requirements of the

course for which these CD-ROMs were being used and this extrinsic factor has implications on their motivation.

STUDENT A: because we were a lot more interested by default because we have an upcoming presentation in a few days so we all wanted to learn like what new can we do our presentations

Justification:

The participants of the focus groups were initially more critical of the first CD-ROM that they worked on which was *Listening to Lectures*, however not so about the *Seminar Skills*. An interesting phenomenon that was observed was that after having given vent to their negativity most of them had a certain reversal of attitude where they started seeing the value of what they learnt. Incidentally, this phenomenon could possibly be explained by how the role of meetings is perceived in Human Resource Management practice (Martin 2000). When an unpleasant move by the company's top brass has to be justified and put in practice departmental meetings are called so that employees give vent to their dissatisfaction and the innovation or recommendation by the administration is allowed to be criticised openly in these meetings and given the impression of being negotiated to foster a sense of inclusion in the decision making process. Advocates of the innovation will rise from amongst the disgruntled group if given sufficient time to get the negativity out. This appears to have happened in the focus groups of this study, as is evident from the quotes included in this section. (Full quotes in Appendix 4.6/table 4.6).

In the immediately following quote student B from focus group one (FG One) is trying to convince others of the fallacy of their demanding longer video clips of the lectures with greater emphasis on note taking, in *Listening to Lectures*.

STUDENT B: But if all of us are complaining about being boring and being dry then if do you expect that the noting exercise is actually 30 minutes long would you be entertained by the lecturer

A grudging reversal of opinion and justification of learning is evidenced in the following comment.

STUDENT L: no but the thing is it teaches you exactly what content is necessary you know ... (Appendix 4.6)

A student realises the value of new learning that has taken place.

STUDENT H: ...CD two on the other hand I mean basically no-one's really taught us how to do a presentation properly ... (Appendix 4.6)

The issue of this learning being without any pressure of impending exams was mentioned in opposition to others' demands of having some assessment to keep them motivated. This point was the salient distinguishing feature between the two focus groups of this study that FG-1 participants wanted assessment to be an integral part of the program but FG-2 participants felt that they were relaxed because there was no pressure of exams. This difference is brought out in debate on the presence of humour and its affective dimensions.

STUDENT J: alright look there's absolutely no pressure on doing that CD you don't have any pressure ... (Cf: Appendix 4.4)

STUDENT J: that's not what I meant it's basically aimed at keeping your attention about something ...

STUDENT H: ...if you're having classes and you have an exam tomorrow and the tutor starts cracking a joke and it's like please get on with it

The following comment from one participant who was sceptical about the materials throughout is an example of justification of the entire learning exercise which this focus group may have achieved through the opportunity of reflection and sharing of their views that it may have presented to the students. This student however had been unrelenting in his criticism yet had already recommended the materials to his brother.

STUDENT C: as a matter of fact I've actually told my brother to get this EASE CD ... (Cf Appendix)

Motivation:

Participants talked about motivation in relation to learning richness and usability the other two themes of this analysis. They also talked about other factors like the opportunity to get admission in a good university if they are fluent and this aspect would be incentive enough for them to work on CD-ROMs like EASE.

STUDENT L: I think the fact that you're going to get yourself into university is motivation enough I seriously don't think you need anymore incentive than that.

Another student wanted to hear the complete lectures from which video clips were taken and felt short-changed by the shortness of the video clips.

STUDENT G: ...it was really sad that we couldn't hear the whole lecture... want to listen to the whole lecture it's just the whole learning they just got chopped off too soon.

However this sentiment was scoffed at as one-off and not representative of the entire group in the following comment:

STUDENT F (F): I mean he just likes the lectures and he wants to go to Warwick end of story ... (Cf. Appendix)

Another feature of *Seminar Skills* was the presence of students that they felt that they could relate to.

STUDENT M(f): you can relate to them I was like I know what it's like to stand up there and not know what to say

4.3 Focus group as an evaluation method

Focus group interviews for the purpose of evaluation of multimedia materials is a method which has flexibility, and gives the opportunity for significant probing of respondents thinking and a high degree of 'psychological depth', defined by Mariampolski as the 'investigation of motivations, associations and explanations behind product preference'(2001:49). This depth of investigation and reflection on learning through this electronic medium was corroborated by the following comment of a participant:

STUDENT F: for our needs it depends on the audience EASE would be brilliant for an audience which is not doesn't have the same perspectives with the learning experiences as us I mean it did help us but by bringing out the subconscious things to our attention and by doing that definitely I'd never thought of many many many things before it I did the program but then these things were I think they were already subconscious and you know

Although this is a negative comment in terms of the materials not being a direct match to the learning needs of the students yet the student exhibits a certain level of maturity as a learner for recognising that the CD-ROM helped them become aware of and externalise certain subconscious academic skills they had not accessed prior to working on the materials.

4.3.1 *Two focus groups*

It was decided that conducting more than one focus group would be a better strategy in order to have grounds of comparison between the two identically designed and conducted focus groups to address the issue of validity and reliability of data collected through this method. Moreover, since the outcome of a single session may not be representative because a single discussion may have focused on a subset of the issues or minor aspects of the system. The two focus groups of this study did show a difference in approach but the evaluative concerns that emerged from the two groups were the same; however the degree of importance given to each theme varied in the two meetings. This can be explained on the basis of the personality differences and the gender mix in the two groups. The first group had more loquacious members who were less inhibited; incidentally this group had only one female participant. The second group had half female members. The second focus group participants came across as less loquacious, this could be attributed to the gender mix of the two groups as girls tend to be culturally retiring and easily intimidated in the presence of more talkative male fellow students. However the interaction was more judiciously spread with regards to air time in the second focus groups among members as compared to the first one where members took long speech turns.

4.3.2 *Role of the moderator*

The moderator in focus groups needs to be skilled in group facilitation and communication to make a focus group successful. It is not as simple as preparing questions, because the moderator needs to facilitate and guide discussion in real time. Although adequate preparation and training had been undertaken, especially since I was already aware of the pitfalls inherent in the role and the role had been practised in the pilot, yet the video of the focus group revealed that my role as moderator was perhaps intrusive at times. The justification of this could be that the discussion could not be allowed to degenerate into an incoherent scuffle where everyone was talking at the same time. I had to intervene with comments like “one at a time please”. At times it may appear that I was participating in the discussion by making suggestions in order to elicit more exciting data. I also appeared to be justifying the materials and the choice of these CD-ROMs for the course. In response to a question on the environment of the CD-ROM a student commented “I actually went to sleep” and I retorted “oh it was because you were tired because of the time of the class”.

In general, teacher evaluator's bias or intrusion can sometimes be explained by the fact that teachers are heavily biased towards one set of materials because they consider them good for their students as ethnocentrism (being immersed in the context) could cause 'insider bias' in sociological research (Schopmeyer and Fisher 1993). In this study, I would agree that the occasional intrusion of my own views in the discussions may be in one sense inappropriate for the 'purity' of the method. In another sense however, I would also argue that had I not intervened, there was a risk that the whole debate or focus group discussion would centre on the question of the relevance of the materials to my students' needs and not focused on the materials at all. The question of relevance was important but had been debated enough and the discussion needed to move forward, therefore I intervened. Tierney and Dilley (2002:465) state that interviewing practices "are not set in methodological stone" and are likely to adapt to changing times and differing contexts. Johnson and Weller (2002:492) in their description of elicitation techniques for interviews state that the "exploratory or emergent character" of these methods "...assume no *a priori* knowledge of informant understandings on the part of the researcher; thus researchers can use them to describe informant responses while minimizing the amount of researcher bias introduced into the research process". I was surprised at the information given by respondents; being familiar with the context had not prepared me as the moderator for the students' views.

The overall impression was that the students were certainly not intimidated by the presence of the teacher researcher and were speaking their mind freely. Despite the fact that the moderator was also their teacher they criticized the course and the materials without inhibition.

4.3.3 Salient features of focus group evaluation

In this section a discussion of focus groups as a data collection method for an evaluation will be discussed with particular reference to evaluating language teaching multimedia materials.

Whitelock (2000:8) discusses the importance of three main parameters to keep in sight when designing evaluation studies 'data size, study cost and richness of evidence'. Selection of a data collection technique will be dependent upon costs of the evaluation from the design phase to the execution and analysis phases. (Almstrum *et al.* 1996, quoted in Whitelock 2000). Applying these parameters to focus groups in this study, the method performed well because it was cost effective, data size was manageable and most importantly evidence was rich in its depth and specificity. The data size was limited in this case but data from 5-6 focus groups could be

collected without impinging on costs too much. Compared to other methods, the costs involved in use of this technique are moderate.

Evidence is rich on a number of counts. First the social dimensions of interacting, discussing and relating to other members of your peer group yields rich and authentic data which is elicited in a relaxed atmosphere. However this conversational feature of focus group made analysis difficult because there was no linear order (for example of one to one interviews). An example of this occurred in FG One where students were asked about the instructional format and sequence of exercises in the CD. After a discussion on sequence, students digressed to consider an "irritating flaw" or a "logical error in the programming of the CD". (TURNS 126-133)

STUDENT B: at times the vocabulary exercises were annoying as well because there were some sophisticated terminologies used in the lecture which obviously at least I didn't know the spellings of

.....

STUDENT C: by the way I figured this out a little bit later if you don't type in an answer for example if press the done button twice it will give you the answer just like that I mean sometimes I didn't realise it was wrong so I just pressed the done button twice and it gave me the answer. (Cf Appendix 4.7 for complete discussion).

This propensity of focus group interaction to digress is corroborated by literature on focus groups which suggests that one of the drawbacks of the data collected through focus groups is that it tends to have low validity and are very difficult to analyse because of their unstructured and free-flowing almost "chaotic nature" (Kitzinger, 1994). Yet at the same time it also throws up a wider range of issues.

According to Merton *et al* (1990) and Morgan (1997) the effectiveness of a focus group can be measured based on four broad criteria: range, specificity, depth and personal context. These criteria from the literature on focus groups also resonate in the overarching criteria given in Chapter Three (section 3.8). My sub-criterion of *productivity* corresponds to 'range' of the data identified here and my criterion of *nature of information* corresponds to 'specificity, depth and personal context' identified in focus group literature. In the following section some of these criteria which are specifically drawn from focus group literature will be applied to the data of the two focus groups of this study.

A student commenting on the importance of being familiar with different accents and pronunciations exhibits personalisation to his given context and shows a range of application of the issue that I had not anticipated, having only taken into account students talking about international accents and not from within the country itself. (Appendix 4.7 FG-1 TURN 17)

STUDENT G: even in an intellectually homogenous university like LUMS where the academics and the instructors and the students are primarily Pakistani you can't escape the fact that some people are from different areas geographically and they've been through different systems so even here we experience a vast difference in accents and pronunciations and something we've had to grapple with coming to Lahore from Karachi so I think it's very very apt even though no material can completely encompass the need for but I mean you need to highlight it it's definitely an issue

Another instance of the range of comments that appeared in the discussion is in the next comment. Student B explains the need to understand the structure and organisation of lectures when various aspects are being debated by group participants. In TURN 81 of FG-1 (Appendix 4.7) a student evaluates EASE and introduces “consciousness raising” as a theme. No one picks it up till TURN 83:

STUDENT B: okay basically I think there are a lot of things that we take for ...I thought lecture organisation a lot of critical points that we tend to overlook were actually we were actually forced to actually look at that you were forced to see where the lecturer gave pauses where the lecturer was laying stress on where he was formal informative whatever different aspects of the lecture were given importance ...I think on a subconscious level we've improved a lot ...

Other members do not pick this point up and continue with negative criticism because they are dwelling on the first half of what is said in TURN 83 and they need to mull over and assimilate the latter half. Another explanation could be that these two group members are perceived by the others as supportive and positive about the experience whereas the other five are eager to voice discontent. The student who originally made the point about ‘consciousness raising’ is the one who picks it up again. (FG 1 TURN 91) similarly STUDENT B continues with the point about lecture organisation he started in TURN 83 earlier.

STUDENT G: I think I can shed some more light on this there's an Eastern saying which says when somebody delivers a speech and a hundred men listen to it each man walks out with his own understanding so basically when you look at it from the lecturers point of view and you try and analyse it and break it down as you hear it you're probably going to walk out closer to what the lecturer's perspective is and that should and the who student walks out on the same page as the lecturer will be the best of like... (91)

STUDENT B... at some point we might think a peripheral point was a main point but if we had noticed the pauses and the organisation the peripheral we might be giving emphasis to a point which a lecturer didn't think was of primary importance so thinking like the lecturer would help us be on the same page as him... (92)

Briefly the aims here were that the combined efforts of the group should produce a wider range of information, insight and ideas. There can be a "snowball" effect with one individual triggering off a chain of responses from the other participants. In the earlier example there is clear evidence that this "snowball" effect did not impede free expression of negative opinion. Positivity had fewer followers and failed to create the "snowball" effect whereas negative comments seemed to have more members opening up and supporting the negativity. This point was borne home by the following exchange in focus group one (Appendix 4.7 FG-1 TURN no: 112-114).

STUDENT F (F): but it becomes a bit monotonous ...

STUDENT E: And there were a lot of slow periods

The word 'monotonous' triggered a chain of negative responses when the group was discussing the sequence of instructional content in the CD. Moreover, the group members found comfort in the fact that their feelings or opinions were not greatly different to that of their peers and they can expose an idea without having to defend or elaborate on it. This evidence is supported by Krueger (1988) who found that focus groups enable individual responses to be spontaneous and provide an accurate picture of their position on a particular issue.

This became more pronounced as the focus group progressed, a few members dominated and the others became quieter because their views were being expressed by proxy by the more eloquent or assertive members of the group. Another reason for some members becoming quiet can be explained on the basis of group formation dynamics. Groups undergo different stages before they become problem solving teams. In literature on human resource management/ organisational behaviour theoretical evidence abounds on the stages of group

development but in common parlance these stages of group behaviour are referred to as: forming, storming, norming, and performing. In other words groups undergo a conflict stage and then a pattern of interaction and group leadership is established and the group starts performing. Even in the short duration of focus group interaction this dynamic would be at work at an elementary level but these students, who had worked together for the duration of the course, had achieved a certain comfort level which could also explain the silence of a couple of members

Focus groups also offered a number of benefits, particularly the potential for group interaction and the creativity fostered by the group dynamic. It was thought that the use of a homogenous peer group who having worked together for six weeks and having achieved a certain comfort level would minimise potential conflict between respondents and counter any 'spiral of silence' or 'stereotype threat' effects (Davis and Silver 2003). Both the 'spiral of silence' and the 'stereotype threat' are derived from a phenomenon known as 'social desirability bias'; where respondents' responses are coloured by their desire to maintain a specific image in the eyes of the interviewer or the other members of the group (Davis and Silver 2003). Glynn and McLeod (1984) identify the 'spiral of silence' as a phenomenon that occurs as part of a group dynamic. Those individuals who notice that their own opinions are either accepted or spreading will voice these opinions self-confidently. In contrast, those who notice their opinions are less well accepted or rejected will be inclined to adopt a more reserved attitude and remain silent.

Davis and Silver (2003) identify that 'stereotype threat' occurs most often amongst those who belong to any stigmatised group and identifies particularly African Americans, women, and students from low socio-economic status. Preconceptions about intellectual ability or competence create a burden as an individual's anxiousness to disconfirm the stereotype, and therefore interfere with their response to the research. There was some danger of this phenomenon happening in the highly achievement driven and grade conscious environment of the data collection context of this study. The status of English as a means of upward mobility and the language of the elite creates certain undercurrents to interactions spoken entirely in English. Superiority is associated with glibness even in native speak but superior class and superior intellect are also associated with fluent expression in English. What was obvious was that students with greater facility of expression were allowed more air time

because the group felt indulgent towards them either because they were in awe or because of having accepted them as leaders.

Focus group interview as a method revealed rich and varied data as is evident from the quotes presented here as sample from data. Participants in the focus group spoke freely in response to each other's comments and many a time took the initiative to disagree and introduce new threads. For instance in FG One TURN 258(follows) the student is disagreeing with other members and giving a surprising twist to the rationalisation for his disagreement with them. He suggests that the working environment of *Listening to Lectures* is realistic because 'boring' and not distracting.

STUDENT E: [?] I actually liked the visual format of the first CD better because it was light colours it was relaxing and yes it was a bit boring but look within LUMS most of our lecturers are boring it's traditionally it's a fact of life most lecturers are a bit dry they're a bit tedious and those are the most challenging to follow so i think you know CD one was a more realistic selection you want to have a realistic selection of lectures

Kitzinger (1994, 1995) argues that interaction is the crucial feature of focus groups because the interaction between participants highlights their view of the world, the language they use about an issue and their values and beliefs about a situation. The language used by the participants clearly sets them apart. The choice of diction and the level of sophistication they were able to employ in self-expression made them feel a sense of elation, which is not present in other means of evaluation. This social dimension of personal expression and feeling good about it is a key stimulant in eliciting rich and honest data through focus group interviews. Other evaluation methods where they were given the opportunity of self expression were the think aloud verbal protocols but it was a linear non interactive expression where the stimulus of others was not present to keep the motivation levels high for continued valuable input on the software. This was evident by the enthusiasm with which they participated in the discussion. It would be wrong to say the enthusiasm was uniform amongst participants but it definitely gave the impression of being infectious. Their views were freely expressed without fear of disapproval from fellow students or intimidation at the presence of the teacher researcher.

According to Kitzinger, (1995) interaction also enables participants to ask questions of each other, as well as to re-evaluate and reconsider their own understandings of their specific experiences. There was clear evidence of

this in both focus groups conducted for this study and students were agreeing and disagreeing with each other and also giving space to each other's difference of opinion.

Another benefit is that focus groups elicit information in a way which allows researchers to find out why an issue is salient, as well as what is salient about it (Morgan 1988). As in the context of this study electronic multimedia materials were used as opposed to traditional materials the depth of discussion and the rationalisation and justification that was going on could be very important for future adoption of e-materials. A case in point is the following student's comment at the end of the focus group when throughout the recorded session he was negative in his criticism of the materials:

STUDENT C: as a matter of fact I've actually told my brother to get this EASE CD ...

Consequently, as Lankshear (1993) points out through this method the gap between what people say and what they do can often be better understood. If multiple understandings and meanings are revealed by participants, multiple explanations of their behaviour and attitudes will be more readily articulated as became evident in the discussion of sub category *Justification* in section 4.2.2.3.3.

The affective dimensions of participating in focus group research came across as students experiencing this sense of importance at being included in the process of decision-making about the value of the teaching materials they had undergone. They took this role seriously and repeatedly digressed to administrative and management issues of using these materials in the context of their needs and suggesting to the researcher how it should be done differently next time. If a group works well, trust develops, and the group may explore solutions to a particular problem as a unit (Kitzinger 1995), rather than as individuals. Not everyone will experience these benefits, as focus groups can also be intimidating at times, especially for inarticulate or shy members.

Although focus group research has some limitations which can be overcome by careful planning and moderating, certain drawbacks are unavoidable and peculiar to this approach. The researcher, or moderator, for example, has less control over the data produced (Morgan 1988) than in either quantitative studies or one-to-one interviewing. The moderator is required to allow participants to interact, ask questions and express doubts and opinions, at times feeling helpless because unable to exert any or little control over the interaction other

than generally keeping participants focused on the topic. By its nature focus group research is open-ended and cannot be entirely predetermined.

Another feature of focus group data that Kitzinger identifies needs to be kept in view is that it should not be assumed that the individuals in a focus group are expressing their own definitive individual point of view. They are expressing an opinion in a given context, which is specific to a culture, is bound by time and occasion therefore has temporal and spatial limitations and so sometimes it may be difficult for the researcher to clearly identify an individual's particular meaning (1995). This has implications for validity and reliability of focus group data.

Another important feature of multi media evaluation that focus group data upholds and delivers on is this aspect of transferability of learning identified by Whitelock (2000:10): *"Future evaluation work also needs to concentrate on the transfer of learning that takes place from the multimedia systems to real life situations."*

Evaluation through focus group interaction with participants reacting to focussed stimuli on the target learning materials and reflecting on it renders this method an effective one to determine whether students' feel they are able to apply this learning in real life. In the context of this study when participants acknowledge developing an inherent understanding of structure and organisation of lectures and applying this understanding to note making in classes then learning has transferred. Other methods could also measure this like tests and TAP but only in focus group would the opportunity to question your own learning through others comments present itself.

4.3.4 *The method's performance on the criteria of evaluation*

The following table shows focus groups performance as an evaluation method measured on the criteria established in Chapter Three, Section 3.8.

Rating Scale:
Negative: (-,-,-) very poor/ (-,-) poor/ (-) slightly poor
Positive: (+) ok/ (++) fair/ +++good/++++very good

The positive and negative symbols denote the degree to which these were good or bad aspects, however the two columns labelled 'Low' and 'High' further elaborate the attributes of the method.

Table 4.10 Performance on Criteria

Criterion	Focus Groups: Positive /Negative	
	Low	High
Cost Effectiveness	Time /Finance (+/+ /+)	Effort (-/-)
Time	Little time required (+/+ /+)	
Effort		Effort in training moderator (-/-)
Finance	Inexpensive method (+/+ /+)	
Ease of Use	Easy to Implement (+)	Preparation and moderator training/Data Preparation (-/-)
Preparation		Moderator training required(-/-)
Implementation	Easy to Implement (+/+)	
Data Preparation	(+ /+ /+) overall impression possible	For research preparation required(-/-)
Bias: Etic: Researcher /Emic: Respondent	With careful training etic bias can be low (+/-)	Emic/Respondent bias high (-)
Ecological Validity: Researcher/Environment Intrusiveness(Hawthorne Effect)		Researcher present physically and recording so high (negative). Environment where peers present provides an opportunity to show off.
Nature of Information Deep/Surface, Subjective/Objective		Deep information being revealed of a higher conceptual but subjective nature (+/+)
Immediacy of Response Immediate/Delayed		Results immediately available for practical purposes without having to be analysed (+)
Usability Measure Satisfaction, Effectiveness, Efficiency	Efficiency :picked at a very basic level wherever there was some obvious error in the system(-)	Satisfaction and Effectiveness picked up better but not as effectively as in other methods.(+ /+)
Robustness of method Validity/ Reliability/Productivity	Reliability may vary due to different cultural and educational contexts (+/-)	Validity high because the method measures what it is supposed to measure(+)

This study found focus group interviews to be a *cost effective* method because the time and financial layout was low and inexpensive. However effort is required in training the moderator of the focus groups. Some *preparation* in terms of the training of the moderator and preparing the materials and questions is necessary and can be time

and effort intensive, although relatively easy to implement. Data from this method does not require any preparation if it is used solely to gain an overall impression of the materials. However, for research purposes or if the evaluator has to prepare a report data needs to be transcribed and prepared for analysis which makes this method not so easy to use.

The teacher evaluator using this method has to be careful not to be misled by the apparently easy implementation of the focus groups. Unbiased and objective moderation is key to using this method for which adequate training measures have to be undertaken. *Researcher* bias has to be kept low and the moderator has to stay objective, which in practice can be difficult to achieve. *Respondent* bias cannot be avoided in this method as the method asks learners for their opinions. They may try to be objective but this cannot be taken for granted.

The *ecological validity* (Preece *et al.* 1994:698) or the influence of the environment and the presence of the teacher /evaluator/researcher/ can influence the nature of data gathered from a method. In focus groups of this study the Hawthorne effect or researcher intrusiveness was present. This intrusiveness can be lessened in well prepared focus groups but cannot be eliminated entirely. Intrusiveness from the environment also contributes to the artificiality of the data gathering situation. The influence of other participants was noticeable; a tendency towards "group think" was observed when conversation was dominated by a forceful speaker. Again the nature of the method is such that peer influence can be lessened but cannot be avoided.

The *nature of information* was mixed. It appears deep and at a higher conceptual level when the participants are trying to explain and justify themselves in an extended debate with their peers, but at other times it appears facetious and shallow when they are trying to impress each other or are thoughtlessly toeing the party line; in other words when they were indulging in "group think". However it is difficult to isolate and verify instances of groupthink and the benefits of focus group interaction outweigh the weaknesses. The interaction and debate in focus group does provide conceptually deep and descriptively rich information in its own right.

The most positive aspect of focus groups was their performance on the criterion of *Immediacy of Response*. Information about the value of a product is immediately available to the evaluator. No tabulation of results or data analysis is required. An overall impression of the materials being evaluated is available directly in a matter of few hours. The difficulty of obtaining an overall impression has been considered a weakness of software

evaluations and methods (Haugland and Shade, 1990; Reiser and Kegelmann, 1994; Wilson, 2001; Scantlebury *et al.* 2001). Focus groups can be an effective method for evaluating not just CALL MM materials but for all learning software. The difficulty may be getting together a homogenous user group with comparable levels of involvement with the software. In an educational setting of a UK university homogenous groups may be somewhat difficult to assemble considering students may be from different departments studying different courses so their needs expectations and expertise with the computer may vary. However in the peculiar ESL language learning situation of my informants the common denominator could be the level of proficiency in English and computer literacy.

As a *usability evaluation measure* focus groups do not deliver on the scale of *efficiency* unless the moderator reminds the participants repeatedly to talk about the usability of the multimedia materials. The participants are keen to talk about emic aspects like their feelings and therefore as a method it elicits *satisfaction* related usability comments and to some extent *effectiveness* comments but the *efficiency* scale should best be measured by other instruments.

This study's use of focus groups as a method measured well on the criterion of *validity* provided the discussion is not allowed to digress extensively and the moderator tactfully keeps bringing it back on track. The *reliability* of focus groups as a method is dependent on the context to a certain extent. The method will elicit data in the same way irrespective of context but it may not be the same kind of data (content or polarity of opinion may differ from context to context). For example, participants from a Chinese context will also argue, debate, indulge in group think but may give a positive verdict on one aspect of the materials or their overall impression of the materials may be good and Canadian focus groups will also argue, debate, and indulge in groupthink but their views may be opposite to those of the Chinese.

If the two groups genuinely hold these opposing views, the method seems reliable. However in a multicultural setting it is important to bear in mind how cultural norms affect behaviour – i.e. whether the Chinese and the Canadians really share the same views, but express them differently because of their cultural backgrounds. In some parts of the world people may be trained from an early age not to express disagreement, while in another part of the world they may be rewarded for being critical. If the moderator is not sensitive to cultural norms this may affect reliability.

The *productivity* of the method is fairly high as the students have the added stimulus of each other and want to impress each other.

In sum, the focus group is a practical, quick and efficient method which elicits adequately rich data to provide an overall impression of MM materials.

4.4 Summary

This chapter has tried to demonstrate the efficacy of focus groups as an evaluation method for CALL materials. This was done by first establishing the kind of questions an evaluator of CALL materials would be interested in finding answers to. The findings of the focus groups interaction were analysed first from the perspective of the teacher researcher trying to arrive at a decision regarding particular materials. Secondly, elicited data was then analysed to show how the key attributes of this method (as identified in the literature) were upheld. Finally the method's performance was assessed on the evaluation criteria especially compiled to answer the research questions of this study.

The nature of qualitative enquiry is explored further in the next chapter which scrutinizes the performance of retrospective protocols as an evaluation method for CALL MM materials.

CHAPTER FIVE: RETROSPECTIVE PROTOCOLS AS AN EVALUATION METHOD

5.1 Introduction

This chapter discusses two variations of retrospective protocols; oral and written, that I used to gather evaluative data. Accordingly, the chapter is divided into two main sections namely 1) Retrospective Oral Protocols (ROP) collected on WAV, and 2) Retrospective Reflective Written Accounts (RRWA) collected through E-mails. A description of the hybrid method of verbal protocols that I used to gather data based on the lessons from the pilot study is given in section 3.6.2 of chapter three. I will discuss the effectiveness of this method from two perspectives, as an evaluator of EASE, and second the researcher, using the evaluation criteria established in chapter three. I will proceed to do the same for RRWA. I will first discuss the effectiveness of the method from the point of view of the evaluator of EASE, followed by an analysis of the effectiveness of this method based on the criteria established in Chapter Three. As the two methods are close there are reasonable grounds for comparison so this chapter will end with a comparison between these two approaches.

5.2 Findings and analysis

The procedure followed for data collection and analyses is given in section 3.7.4 of Chapter Three. I collected copious amounts of data by recording oral protocols on WAV and receiving reflective written accounts by email: over 400 written and 500 oral retrospective protocol accounts. In this section I will discuss the codes that emerged from an analysis of the data and the conclusions that can be drawn from the analysis.

5.2.1 *Evaluative comments from oral protocols*

The themes and categories that emerged in the coding of focus group data are taken as a starting point for the themes and coding strategy for ROP and RRWA. Table 4.6 from Chapter Four was adapted to create Table 5.1 below.

Table 5.1: Themes and Categories

Themes	Categories	Sub Categories	Definition
Learning Richness	Appropriateness		Relevance of materials to the needs of students and the requirements of the course
	Content	Difficulty level	Imbalance in level, and treatment or sequence of instructional content.
		Redundancy	Content that was considered repetitive or unnecessary and the mismatch between contents and individual expectations
		Sequence	Order in which the units/exercises occurred
		Exercises	Specific commentary on the nature of exercises
		Clarity	Comments on the clarity or lack of clarity
	Expectations		What the students' hoped to learn from the materials and how this was going to be useful in immediate or long term future academic or professional careers.
	Reflection		Evaluative commentary on students' own performance or learning
	Motivation		Self evaluation assessment mechanisms like scores to assess progress, and authentic content of video clips motivating
Improving Competence	Usability (the functional and ergonomic aspects of the CD-ROMs or HCI factors).	Glitches	Programming errors or system failure that the students discovered
		Satisfaction	Expression of satisfaction (emic) at usability features of CD
		Effectiveness	Effectiveness of functional features(etic) of the CD navigation, online help, adaptability.
	Suggestions	Suggestions for improving the program's functional and pedagogical competence.	
	CD-ROM Comparison	EASE <i>Listening to Lectures</i> and <i>Seminar Skills</i> are compared with each other.	
	Motivation	Features in design that affect motivation	
Affective Dimensions	Management	Management of learning innovation	
	Justification	Rationalisation of earlier criticism students have tried to rationalize their earlier impressions and come to a compromise or are themselves providing a justification of what the program aimed to achieve.	
	Motivation	Motivational and affective dimensions of authentic cultural exposure over all perceived benefits or drawbacks	

This table is based on the hierarchical thematic coding of categories that the program NVivo permits when creating "trees/nodes and children". The general overarching theme of the experience of learning was *Learning Richness*. The sub themes that emerged were *Appropriateness*, *Content*, *Expectations*, *Reflection*, and *Motivation*. Of these *Expectations* and *Reflection* were new categories and the category which was considerably expanded into further sub categories was that of *Content*. The new categories in *Content* were coded as *Difficulty Level*, *Redundancy*, *Sequence*, *Exercises*, and *Clarity*. The evaluation method encouraged specific and topical discussion, so it was also necessary to expand the subcategory *Usability* under the main theme of *Improving Competence* to create three further subcategories: *Glitches*, *Satisfaction* and *Effectiveness*. A brief explanation of each category is attempted in Table 5.1.

It was difficult to create codes and categories which were entirely mutually exclusive; therefore only the focal point of each comment has been used for categorisation, which means that the same comment may be coded as belonging to more than one theme and may be coded under two or more different nodes. The following comment illustrates how this comment could be coded at the subcategories of both *Exercises* and *Difficulty Level* in the main category of *Content*.

58: later on in this unit they bring up vocabulary exercises again which initially seemed a bit simple and easy but exercise 45 on positive and negative connotations was quite hard... (WAV Session3 Node Difficulty level, Student RP11F)

The following comment was double coded under *Expectation* and *Reflection* as it is both a reflective comment ("it was a constructive experience") and an expression of future expectations ("helpful for us in future courses").

5: On a final note I believe that it was a constructive experience working on the EASE software and I hope it can be helpful for us in future courses. (WAV session 10-12 CD2/5. Node Expectations, Student RP15M).

The numbering of each coded comment that appears here and in all of Chapter Five is sequential according to NVivo. The program NVivo has its own unique system of numbering which has been kept here for ease of reference to the coding reports and comments in Appendix 5.1.

Any teacher/researcher evaluating EASE for his/her students would be interested in a *at a glance* interpretation of the data to determine whether the students liked the MM program or not.

Nine out of 12 sessions of the WAV protocols contained comments coded within the category of *Appropriateness*; comments coded in this category tended to express either clearly positive or clearly negative attitudes. Within this category in NVivo further coding into negative or positive comments was possible. I counted all the positive and negative comments in this category which are presented in Table 5.2.

Table 5.2 Positive and Negative Comments

Categories	Codes	Positive Comments WAV Sessions 1-9									Negative comments WAV Sessions 1-9								
		1	2	3	4	5	6	7/8	9	Total	1	2	3	4	5	6	7/8	9	Total
Appropriateness		23	41	29	36	40	45	84	16	314	2	15	7	4	11	6	7	0	52

This counting exercise although simple was nevertheless time consuming and may have implications for the effectiveness of the evaluation method in terms of the criteria of *Cost effectiveness* and *Ease of Use*. However, it illustrates one simple method of data analysis that a teacher, researcher or evaluator of EASE may undertake. It revealed that the majority of comments from the 45 plus students working on the two EASE CD-ROMs were positive. For the *Positive Comments* category a total of 314 comments from nine WAV recorded sessions and another 5 comments from residual WAV files which were recorded separately therefore a total of 319 positive comments as opposed to 54 in *Negative Comments* (52 from first nine sessions plus 2 from the residual files). The students worked on the first CD-ROM until approximately session five. From session six onwards the number of negative comments reduces further and the number of positive comments increases. This is on account of the students commenting that they liked the *Seminar Skills* more than the *Listening to Lectures*. They felt that they gained more from *Seminar Skills* and compared to *Listening to Lectures* it was more *appropriate* to their needs. A more detailed and more qualitative discussion of this category is given in section 5.2.1.1. below.

The other categories could also be analysed according to this simple negative/positive formula but this analysis is not undertaken here as it would yield similar results and for the purposes of this study other themes need to be discussed qualitatively. A discussion of all the emergent themes from the perspective of an evaluation of EASE follows. All sessions in which the students worked on EASE were recorded and are called WAV sessions here. The students worked at their own pace so the WAV session number does not necessarily correspond to the number of the program unit they are working on. WAV session five, for example, may include some students working on Unit 3 of the first CD, *Listening to Lectures*, others on Unit 5 and yet others on Unit 1 or 2 of the second CD, *Seminar Skills*.

5.2.1.1 *Learning richness*

All comments which referred to the quality of the learning experience and the relevance of the program to student needs are included in the main category of *Learning Richness*. Only the first category of *Appropriateness* is

discussed at length to convey an overall impression of the nature of the comments. A few comments from other categories are also discussed but full comments are given in table format in Appendix 5.1.

5.2.1.1.1 *Appropriateness*

Appropriateness pertains to those comments which discuss the relevance of the program to the students' individual needs and the requirements of the course. Both negative and positive comments show that individual students were evaluating the efficacy of the materials in relation to their individual learning needs. I am going to first look at positive comments related to this theme and then negative.

POSITIVE COMMENTS

The first ROP session recorded on WAV contains mostly positive comments in relation to the first CD-ROM *Listening to Lectures*. The students match the aims of the program to the requirements of their course. The following excerpts show approbation and appreciation together with some understanding of how the program can help students.

4: I think the introduction of EASE was very relevant and the use of both writing that is text and listening will really help me ... (Session One, Node Appropriateness, Student RP 1M)

The students match their perceived needs with the content of the CD-ROM and are hopeful about its efficacy. The following comment shows that the student hopes to gain confidence in public speaking in English.

30: ...overall EASE appears to be a good attempt which I think will help most students overcome their awe for the English language and public speaking... (Session One, Node Appropriateness Student RP 8M)

Yet another student appreciates the inclusion of students in the video clips as a motivating factor as she feels that students learn from peers and their interaction with the teacher and amongst themselves.

178: The best part about this CD is that it's got viewpoints from students as well as lecturers you know we tend to relate to people from our age better ... (WAV Session one. RP20F)

Another general comment about the efficacy of the *Seminar Skills* CD-ROM is based on the advice that it gives:

38: ...we were first made to distinguish between seminars and lectures ... this CD doesn't just make you learn through exercises it provides you with clear and vital information and advice ... (WAV Session 7/8 Student RP30M)

Along with a general commentary on the EASE materials and their aims and objectives as perceived by my students there are also more content and exercise specific comments such as the following:

26: ...the exercises like the one taking notes where we have to take notes those were pretty helpful ... (Session One, Node Appropriateness, Student RP2M)

42: ...the note taking was a very interesting aspect of the whole thing because none of us knows how to take proper notes and listen to a lecture and make notes at the same time... (Session One, Node Appropriateness. Student RP20F)

Content and exercise specific comments show the students' level of involvement with the exercises. In *Listening to Lectures*, the note taking exercises were considered particularly helpful as opposed to the vocabulary exercises which got mixed reviews. The grammar exercises annoyed most students as they were perceived to be below their level; these were only appreciated by a few students, as a refresher course in grammar.

As the students worked through the program their involvement deepened and this is evidenced in the detailed and involved nature of their recorded comments.

30:...with the help of EASE I am learning a lot about evaluation of instructors their attitudes their styles I also learnt about the difference between figurative and literal meanings ... today I also learnt when an instructor is making an argument and I also came to know about two types of argumentation which are thesis antithesis and synthesis I also learnt how arguments are structured ... (WAV session 7/8, Student RP7F)

There is evidence in the following excerpt of an awareness of what has been learnt and an assessment of the student's performance.

46: ... I'm able to evaluate whether the lecturer's claim is strong or less weak now I can easily differentiate in lecture style the manner whether he is serious or comical now I can make judgments about the lecturer's ... I know the literal and figurative meanings as well of the lecture I feel more confident about distinguishing between the significant and the less significant ... (WAV Session 7/8 [Cd1] Appropriateness, RP47M).

The following quote shows a clever student pushing himself further by improvising/adapting on the intended use of the materials. The REWIND option is available in the materials but s/he chose not to use it to make the lecture more challenging and closer to real life.

50: ... the speaker was going relatively fast and I had to train myself to go at his pace and not repeat the lecture again and again because I can't stop press stop and rewind in class... (WAV Session 5 [Cd1] Student RP38F)

Yet another comment shows that students also make comparisons between units, making an earlier unit the comparison point for a later unit. The range of information revealed by this method therefore is broad, deep and discrete.

78: I think Unit Three was probably the most practical and the most helpful of the units so far because unlike Unit One and Unit Two which were quite general Unit Four or Unit Three was specific ... (WAV session 4 [Cd1], Student RP25M)

The designers' decision to use video for listening as opposed to just audio is justified by comments like the following:

58: ...judging the tone of the lecture makes a whole world of difference because occasionally they are used in another context of meaning ... I shall be able to concentrate more deeply and analyse not only the verbal account but also the speaker's body language facial expressions and tone... (WAV session 2 [CD1] Appropriateness, RP8M)

As became evident in the counting of positive and negative comments in section 5.2.1, the students' evaluation of the second EASE CD-ROM *Seminar Skills* received more positive appreciation on account of it being perceived as more relevant to their needs. Comments in the later sessions tend to concern the *Seminar Skills* and

comparisons between the two CD-ROMs are inevitable, with the first CD-ROM *Listening to Lectures* acting as a point of departure.

6: ... this unit was very helpful as it also provided information about mistakes students make during their presentations after looking at those mistakes I can now think of the ways to avoid those mistakes in the presentations which count a lot in my academic career it also gave me a view about the advantages and disadvantages of using the computers and presentation software... (WAV session 9 CD2/Unit 3 Node Appropriateness Student RP14M)

As students moved towards the end of the program their views about their learning grew more positive to the point of recommending that other universities should use EASE to teach similar courses.

297:...the entire EASE CD was really innovating(sic) and the interface used was user friendly I think EASE and other multimedia electronical(sic) education programs are the future of learning and teaching and of course success my vocabulary my communication skills my writing skills the note taking skills are really have been really enhanced after using the EASE I think LUMS and other universities must incorporate EASE and other related programs (WAV session 7/8 CD1/5 Node App: Student RP24M).

Yet another student's comments show the detailed nature of the analysis of the contents of the program. This student candidly admits to not knowing the difference between a lecture and a seminar.

329: Well this is a pretty comprehensive unit because apart from the examples of presentations there are suggestions from the tutors the differences between a good and a bad presentation ... presentations which are meant to be for academic purposes or otherwise the role of tutor and student was very clearly broad ... moreover the difference between the lecture and the seminar was very good because before that I used to think of a seminar as the same thing as a lecture. (WAV Session 7/8 CD 2/1 Student RP 37M Node App.)

The above comments are a small sample of the positive comments relating to appropriateness made by students

NEGATIVE COMMENTS

Units or exercises that seemed inappropriate in terms of level of difficulty, and which did not challenge some of the students, or perhaps they found difficult, received some negative comments. A sample of these is given below.

30...the whole exercise recognising accents was totally baseless and pointless... (WAV Session 2 [CD1/2]

Node Appropriateness Student RP 7F)

As Table 5.2 above indicates, the negative comments are fewer in number than the positive ones yet stronger more emphatic language is used in some instances. Negative comments are also more likely to be accompanied by a reasoned explanation. This is in accordance with politeness theory. The negative comment is the dispreferred response, and therefore requires the use of politeness strategies, such as explanations. Noticing of negative comments or placing any emphasis on negative comments could also be a researcher's bias reaction hence tantamount to intrusiveness and will be discussed at greater length in section 5.4.

30: I didn't find any purpose in having the exercise about backward and forward markers because I don't think it helps us concentrating on helps us in improving our concentration powers...(WAV session 2 [CD1/2 Node app Student RP7F).

In the preceding and following comments there appears to be a certain level of anger and resentment perhaps born out of frustration because the students are unable to see the relevance of the materials to their learning needs.

98: I don't get how we as students are supposed to apply this in a lecture room okay fine we know what transitional statements are but how will it help us in a lecture okay we know the teacher's going on to another point ... but how does that improve our listening skills and the ability to concentrate more on lectures (WAV session 2 CD1/2 Node App Student RP20F)

In some instances the comments question the very inclusion of a topic and its relevance to any context.

248:...the vocabulary exercises were far too easy I didn't understand the purpose of the vocabulary exercises in which we have to list down the nouns adjectives et cetera ... and the exercises relating to figuring out which countries the professors came from were quite irrelevant and stupid discovering which country a person is from by paying attention to his accent doesn't improve one's listening skills ... (WAV Session 2 [CD1/2] student RP46 M)

The role of repetition as reinforcing learning is lost on the following student who perceives it as a negative attribute.

70: The second unit I've completed as well I think it was well constructed and the content of language was good but according to me it was overly repetitive ... it was boring in the end... (WAV session 7/8 Appropriateness Student RP 17M Cd2/2)

Exercises that are perceived to be too easy and insufficiently challenging are considered obvious or pointless just as basic grammar is considered pointless by some.

11:...the exercise on separating literal and figurative phrases was also an easy one I mean it was pretty apparent whether the instructor actually meant what he was saying or where he was not apart from this the exercises in which we were supposed to identify the nouns and adjectives derived from the verbs and their meanings were totally pointless and so were the exercises that required us to fill missing prepositions ... the exercise on intensifiers although it was easy but more or less it was pointless I think intensifiers are very commonly used and sometimes even unintentionally unless of course instructors take special care before uttering an intensifier. (WAV session 7/8 CD1/unit 4/5 Node Appropriateness Para11 Student RP7F)

Impatience at certain exercises which are either not understood or misunderstood is evident in the following dismissive comment.

58: In the vocabulary section of Unit Three I didn't get the point of one of the exercises where you had to see the lecture and see where the lecture uses alphabets from the English language maybe there should be little notes at the beginning or end of each exercise that explained what it is aiming to achieve because frankly I didn't see the point of this one exercise ... (WAV Session FOUR [Cd1] paragraph 58 Node Appropriateness Student RP20F)

The student comment above is also making a suggestion that each exercise should be preceded by an explanation of what the exercise hopes to achieve. It can thus be surmised that even criticism is peppered with suggestions and ideas for improvement as also in the following comment.

5:...but I feel if that was the sole purpose to teach us how to dissect arguments then its not, I don't think its focused enough, I don't think it comprehensive enough, ... the methodology of arguments, how to make an argument how to break an argument and how to analyze and argument how to summarize it, ... but its not as comprehensive as maybe it could have been because the arguments are very central to a student's life ... (CD 1 Unit 6 WAV Residual files. Node App Negative Student RP 52 M)

This initial discussion of positive and negative comments is representative of the polarized views of students on other aspects of the EASE materials. This category was the largest in terms of the number of passages coded under this node therefore a full discussion is undertaken here. Other categories and codes are presented in summary form with few sample comments. More comments from these categories are presented in Appendices 5.1/5.2/5.3.

5.2.1.1.2 Content, expectations, reflection and motivation

A significant number of the comments pertained to the pedagogical content of the units and the nature of the exercises in these units. Selected complete comments are given in Table 5.1 in Appendix 5.1.

In this main category of *Learning Richness* were *Appropriateness* (which is dealt with at length above) *Content*, *Expectations*, *Reflection and Motivation*. The category of *Content* had the most number of comments coded resulting in further sub categories of *Difficulty Level*, *Redundancy*, *Sequence*, *Exercises* and *Clarity*. The codes of *Expectation*, *Reflection* and *Motivation* had fewer comments with *Motivation* related to content having the most mention (amongst these three) in students' comments.

The impression distinctly conveyed is that students would like to be challenged and mentally stimulated. The features of the materials the students seemed to have liked most are the ones they found most challenging. As the students progressed through the units, few students continued to find the exercises insufficiently challenging, whereas others expressed surprise at the 'raising of the bar'. (*WAV Session 2 Para 58 Student RP 8M Cf. Table 5.1 Appendix 5.1*)

One key objective of *Listening to Lectures* was to familiarize the students with different accents that they may come across in the multicultural environment of a British University and student RP 14 M's justifies it.

14: I find it difficult to identify the difference between the speakers that from where he or she belongs to...

(WAV Session3 Para 14 Node Difficulty Level Student RP14M)

Student comments also become much more specific to the pedagogical content of each unit as they progressed through the materials. It would be safe to surmise that this depicts a deeper level of engagement with the program and the method's ability to capture this depth.

The comments continue to be both negative and positive based on the student's attitude and ability levels. The specificity of comments highlights some of the possible difficulties students may face because they are still discovering certain aspects of the program through trial and error.

46: I started with Unit Four I could complete the reference portion easily in a single attempt although it took me like three attempts to guess the number of people mentioned naming the people was a bit challenging because I was getting the spellings incorrect most of the time...(WAV Session 4 CD1/4Student RP3F)

A repetitive strain in the comments is the sense of elation and accomplishment when difficulty level of exercises increases.

155: So far Unit Five is proving to be the most difficult ... analysing the level of conviction and ranking them was even more of a challenge and overall I saw it as a good practice (WAV session 6 Student RP11F).(for Full Comment Appendix 5.1 Table 5.1)

What also becomes evident through this analysis is the differing pace at which students worked through the program. Some students worked much faster than others and the materials accommodated such variation in learning pace. Differences in pace are clearly evident in the recordings for WAV session five as some students talk about unit four and some about unit six, making it difficult to draw generalized conclusions from this session.

5.2.1.2 *Improving competence*

This section deals with the functional and HCI factors of the program more than the pedagogical content. However, my students were not completely able to isolate the two with the result that the discussion on the ergonomic functionality of the program is limited. The codes under this tree node are *Usability*, *Suggestions*, *CD-ROM Comparison* and *Motivation*. (Appendix 5.2 Table 5.2 gives sample comments from each category). The category of *Motivation* in this main category codes those comments that relate to the functional and ergonomic features of the program.

Usability aspect related comments were divided into the sub categories of *Glitches*, *Satisfaction* and *Effectiveness*. Comments in *Glitches* distinguished between functional errors and errors of interpretation by the programmer and suggest that a closer coordination between the materials designer and the programmer could resolve these errors. This category is relatively significant because it can actually help the designers improve the program although the comments may be of little interest to the teacher researcher conducting this evaluation. But this would depend on how many problems the students reported. If there were a lot that might be a good indication not to purchase the materials. However the researcher would be advised to check that the criticisms are fair - in some cases students may have misread the rubric, for example. The sub category of *Satisfaction* had comments pertaining to feelings of satisfaction at having worked through the CD. A sense of accomplishment or achievement is the hallmark of comments coded under this node. Not many comments are coded under this node because this code pertains only to usability-related satisfaction. Comments that referred to ease of use in relation to the usability of the program are coded under *Effectiveness* category.

This category has comments which give *Suggestions* about improving the competency of the program not just in terms of the technical functional aspects, but also the pedagogical content of the program. Students suggested that answers to the exercises should have had video explanations rather than just text and that each exercise should be preceded by an explanation of the aims and objectives that the exercise is hoping to achieve (comment 54: *Suggestion*. Student RP25M. Appendix 5.2: Table 5.2).

Students also compared the two CD-ROMs a student's (RP29M Comment 321: Table 5.2 in Appendix 5.2) highlights the similarity of layout and another one found the CD-ROM easy to navigate because the designers of the second CD-ROM in the series had considered continuity issues and it was motivating for him,

(Comment 333 Table 5.2 Motivation, Student: RP7F). Comments which affected motivational levels either negatively or positively were coded under the *Motivation* node. This motivation section pertains to the CD's functional competence rather than the pedagogical content.

5.2.1.3 *Affective dimensions*

This node coded comments which related to the overall effectiveness of the CD-ROM and how it would affect the students' life and learning extraneous to the CD-ROM itself. How much they enjoyed working with the CD-ROM and what benefits this learning will accrue for them are also coded here. The children nodes under this tree node are *Management*, *Justification* and *Motivation*.

Management has comments which discuss the inclusion of the MM materials in the course and a student suggests that the University should integrate the materials into future communication classes. The positive comments about the materials (comment 118 Table 5.3 Appendix 5.3) are equalized by negative one about the way the classes were organized (comment 93, Table 5.3 Appendix 5.3), showing the diversity of opinion that a good sample size provides. This diversity and the sheer number of specific comments contribute to the rigour and reliability of the evaluation method.

Under the *Justification* node came comments in which students tried to rationalize their earlier impressions and come to a compromise, or provided a justification of what the program aimed to achieve. RP22M (comment 46, Table 5.3 Appendix) makes a valid point about the authenticity of the lecture based exercises but the expression that the program "lacks" is somewhat inappropriate maybe facetious. Student RP13M's comment (74 Table 5.3 Appendix) justifies the premise of *Listening to Lectures* and its inclusion in the course for my students' language skills development. The sequence followed on the course is commented on by Student RP6M (comment 82 Appendix 5.3: Table 5.3) who justifies first going through *Listening to Lectures* then doing *Seminar Skills*. Student comments also give the ultimate justification of saying that it was time well spent.

This category of *Motivation* is different from the other two motivation categories which were in the thematic categories and NVivo tree nodes of *Learning Richness* and *Improving Competence*. This node codes comments which go beyond the EASE program and catalogue the program's possible influence and impact on the students' lives

and how it may or may not help them achieve their larger aims and goals in their academic and professional lives. How happy, satisfied and hopeful the students feel is also mentioned in these comments.

Students saw value in learning ‘to infer meaning of the foreign words’ as good (RP37M Comment 10 Appendix Table 5.3) and felt that their world view has been enhanced by working on this program.

The students expressed appreciation of the exposure to British academic culture in anticipation of going abroad to study (comment 131 Appendix Table 5.3). Comments showed students engaging with specific content and attributing intellectual and creative engagement to the program and how this affected their motivation (Comment 226 Appendix Table 5.3 *Motivation*).

226: Thesis antithesis and synthesis are three terms I have absolutely memorised by the end of Unit 6 ... the synthesis antithesis and thesis exercises allowed us at least me to use my own creativity to use my own intellect to come to a conclusion and not only rely on the assumptions or the conclusions drawn by the instructor (WAV session 7/8 Student RP 43M).

Students also expressed appreciation of yet another affective dimension of coping with the multicultural aspect of academic life (Comment 6 Appendix Table 5.3 Student RP9F).

6: ...a few exercises were about discrimination that we face in nowadays societies and on small scales and large scales basically that was pretty good it gave me the chance to express my views and talk about it... (WAV session 7/8 Student RP9F)

5.2.2 *Evaluative comments from e-mails*

This section reports on the analyses of the 400 plus emails sent by the students. Each of these emails was written after the students had worked for 2.5 hours on a section of the program. Each batch of comments refers to more or less the same issues, as the emails relate to the work done that day in class. However, the written medium seemed to encourage more thoughtful and carefully prepared compositions whereas the verbal protocols which were recorded on WAV seemed more spontaneous. The emails were coded under the same categories and nodes as the verbal protocols yet the nature of the emails is somewhat different. The most obvious difference is the level of formality. Most of the emails maintain completeness of form and structure, even when they give the impression of being hurried or perfunctory (see for example Emails 317 as opposed to

Email 280 Appendix 5.1). (Table 5.4 in Appendix 5.4 gives complete excerpts from emails coded under different categories).

5.2.2.1 *Learning richness*

The emailed comments expressed many of the ideas already identified in the earlier verbal protocols; they tended to be more reflective but contained few entirely fresh insights. They were less spontaneous and more objective than the oral protocols.

Without resorting to counting the negative and positive comments coded under *Appropriateness* it was evident that as in the same category for ROP, there are many more positive comments than negative (although negative comments tend to be more emphatically stated).

*The first class of Communication Skills, in which we were required to use the soft EASE, was if I may say so without sounding corny, was a breath of fresh air. It was different from the mundane lectures I am used to .
...further use of this program will allow me to hone my note taking abilities... (Email 1 Sender RP11F).*

Comments related to the specific content of each unit of the two CD-ROMs were placed in the category of *Content*. Subcategories of this node were *Difficulty level*, *Redundancy*, *Sequences*, *Exercises* and *Clarity*.

*I had a lot of trouble with the identification of functions chapter. It was pretty hard and throughout the exercise
... I came to the conclusion that it was not a very useful exercise. (CD 1 Unit 4 Email 333 Student RP38F).*

Comments about repetition of content, or unnecessary or irrelevant content were placed in the subcategory of *Redundancy*.

Unfortunately, there was also some very easy and irrelevant stuff there. There were too many "mindless" questions, where you had to fill out the exact phrase used by the speaker, word for word. Almost as pointless were the exercises on vocabulary and grammar. (Email 110 sender RP4M).

Many comments show students' awareness of the importance of the right pace of progression and smooth transition between units.

...the structure is very logical as it raises the level gradually and quite properly. (Email 147 Sender RP29M).

Students wrote about their expectations from the program and how certain exercises kept them more motivated than others.

I expect to learn how to be a better student at the university altogether. The cd's are expected to be a source of help to me in case I want to continue my studies abroad. (Email 11 Sender RP22M).

This session was rather stimulating and kept me awake throughout. It was comprehensive and coherent. (Email 100 Sender RP41F).

5.2.2.2 Improving competence

The errors that were discovered by the students and their justifications for these are coded in this node. These comments mirror the ROP commentary with the difference that some element of reflection and a toning down of exasperation is evident, probably due to the relatively delayed nature of this retrospective account.

Then there were some exercises where i got my answers wrong because i entered "We have" instead of "We've".

...So the software fails at such places. (email 131 Sender RP13M).

A repeated concern, also expressed in the ROP oral protocols, was about the functionality of the scrapbook feature of the program. Students who seem to have not had a bad experience with the scrap book are ones who may have learnt to negotiate it. There is also the possibility that some students may not have engaged with the software as thoroughly as others, or some individual machines may be playing up or some students just worked on the scrapbook as was needed for it to work efficiently while others who complained about the scrapbook had not been able to hit the right formula.

...however saving the scrap book is more difficult in this. ive managed to gte(sic) my scrap book lost twice in this session. (Email 348 Sender RP10F).

A direct measure of satisfaction could be the students' sense of accomplishment which is gathered indirectly through this method of reflection as opposed to asking them a direct question such as in a questionnaire makes this method less intrusive. A successful feature of the program appears to have been the dictionary help that was provided in the software and few students did comment on it.

Suggestions to improve the program are also coded under this node and some interesting observations which may be of value to subsequent revisions of the program. The points are significant but subjective in nature. A very practical suggestion for designers not necessarily the teacher evaluator is given about a 'bookmark'. (251 Sender RP13M Appendix 5.4 Table 5.4 Suggestions). A recurring concern of my students was with assessment of their performance and many suggested having an assessment element as a suggestion for improvement. (Email 365 Sender R43M Appendix 5.1 Table 5.6B Suggestions)

5.2.2.3 *Affective dimensions*

This section coded comments which had criticized the way the course, in which they worked on this program, was managed or designed. In the *Justification* category the comments try to justify either a recurrent criticism of the CD-ROM that the students have been making or an earlier negative opinion which has been revisited towards the end. These remarks show reflection on their own attitudes to learning through this program and are constructively self critical. The positive reflections often balanced the negative opinions. As there were detractors of the program/course there were those too who considered it an opportunity to learn through this system

A comparison between the two CD-ROMs of the program is a recurring theme; most remarks validate the use of EASE CD-ROM 1 *Listening to Lectures* prior to EASE CD-ROM2 *Seminar Skills* (Email 291 Sender RP10F See Appendix 5.4 Table 5.4). The students do understand the overall global benefits that working through this program may have provided. The value of this innovative approach of using authentic listening texts used in this program is not lost on the students. Innovative teaching methods are being compared to the conventional classroom in some comments showing that students are willing to learn by new electronic media (Email 12 RP44M Appendix 5.4 Table 5.4).

The specific nature of comments reveals the depth of the students understanding about teaching and learning best practice as is illustrated by the comment where a student points out how exemplifying is important for teachers (Email 62 Sender RP 20F Appendix 5.4 Table Motivation). The value of showing authentic student -student interaction in *Seminar Skills* is validated as learning from peers and relating with them is given importance. (Email 323 RP20F. Appendix 5.4: Table Motivation). My students' preoccupation with assessment and grades persisted

and they said inclusion of assessment would have been motivating for them. Humour was appreciated at all levels in the two CDs. The theme of *justification* revealed some interesting views on the experience.

Why i might have felt more favorably to this CD, in retrospect, i feel is because maybe my listening skills and surviving boring lectures has improved. the first CD actually helps you to go through this one. (378 RP10F).

This concludes the analyses section on RRWA.

5.3 Main findings of the evaluation

The qualitative analysis of the commentary reveals that the majority of learners liked the EASE program. The majority of the students considered the program appropriate for their learning needs and saw benefit accruing to them. They gave detailed feedback on the exercises they liked and the ones they disliked. Note taking though considered challenging was an all round favourite because of perceived benefits. Vocabulary was the least respected as it was considered a distraction from the more serious content such as argumentation and identifying thesis and antithesis.

It was observed that the cognitive load of the program was considered manageable except when the scrap book and other minor programming errors impinged on the smooth interactivity of the program. The category of *Glitches* coded the system errors or snags that my students discovered and which affected the smooth functionality of the program. Some of these errors are deeply embedded and only became apparent after a prolonged engagement with the program across a broad range of users. For this reason this kind of summative evaluation would also be useful for formative purposes to improve the program at its next revision.

The students found the program easy to use and commented on smooth transitions between units and between the two CDs. Integration of different media was appreciated, particularly the use of authentic videos of actual classrooms showing student interaction. There were fewer comments on user interface or usability aspects of the program, perhaps this method or my students use of it did not reveal many usability issues. The other conclusion that can be drawn is that it is well programmed software so does not have too many errors. The students also reflected on their learning process constructively. Approbation and appropriateness of the program was the overriding impression that the analyses upheld.

The learners also gave suggestions regarding aspects of the program that they felt dissatisfied with. For example, a few students commented on the disparity between the difficulty level of the lectures and the exercises that followed the lecture. Certain versions of the correct answer minus the vocal filler expressions like “err/umm” used by the lecturers were not given as correct answers which exasperated some students. The scrapbook could not be saved in CD1 *Listening to Lectures*. Overall there were fewer negative comments, validating a positive verdict on the program.

However this software evaluation is incidental to the main concerns of this study so the next section will evaluate the Retrospective Protocol’s performance as a method of Evaluation for such MM programs.

5.4 Retrospective protocols as evaluation method

The above cited comments serve a dual purpose firstly they document what a teacher evaluator would be interested in finding out about the program and secondly they provide evidence of the method’s performance. Summarily, the qualitative analyses of students’ comments revealed that the method captures the following:

1. *A whole range of diverse opinion*
2. *Deep specific commentary on program content*
3. *Frank and uninhibited commentary*
4. *Students working at a regular pace and sincerely engaged*
5. *Students understand the concerns of the evaluator*
6. *A very learner- centred method*
7. *Greater awareness of learning*
8. *ROP spontaneous in nature and RRWA reflective and tempered*

A discussion of the above findings follows:

A whole range of diverse opinion: Students expressed very diverse opinions but it was possible to cluster them to code them. Had there been too much diversity it may not have been possible to code the comments. The number of respondents in this research and the use of WAV recording software/ emails made it possible to catch this diversity of opinion. However this generalization can not be applied to another research design where there are fewer respondents. Therefore it may not be a function of retrospective protocols as a method but a feature of this study's version of RP. The next point of eliciting deep specific commentary can be the strength of RP and not just the use of the method in this study.

Deep specific commentary on program content: Deep specific commentary emerged as the students worked through each exercise of each unit giving their opinions and evaluating the content on the criteria of their needs and their motivations.

The commentary on the Wav files (ROP) by its very nature was topical and although students talked about what they did correctly and what went wrong they tend to focus more on the problems and difficulties (this could also be an instance of researcher bias as I may have attached greater significance to comments that did this rather than comments that just documented successes) whereas in the emails their reflective comments shift to a discussion of their own performance and become more reflective. One good side benefit about talking about their work in a positive sense even under the compulsion of showing themselves in good light makes them think of the whole learning experience positively.

27: I've started Unit Six now there was a section in Unit Five by the name Attitude we had to perceive and understand what the lecturer was trying to say if he was being sarcastic or jokey and stuff it reminded me of the time when I was doing my O-levels and a teacher I think she was my Geography teacher and she was pretty much against [?] she said [?] and I took it's literal meaning and kept thinking about how that could be and that was stupid there were other exercises that made me realise how bad I am with prepositions ... (WAV Session 6 CD1/ Student RP9F).

The above comment shows a student relating new learning with old seeing connections and the schema being engaged. This is good evidence of the method capturing evidence of learning effectiveness and the mind

processes of short term memory and long term memory in learning through MM application. The spontaneity associated with TAP is also evident here in this ROP.

Frank and uninhibited commentary: The open criticism of all aspects of the program suggests that the students were frank and open in their assessment of the program. It appears that nothing was stopping them from being critical as in the case of focus groups the role of the moderator can possibly influence free exchange of opinion. This is especially important as in the academic culture of the East it is common for students to not openly express criticism of any aspect of a course. However, because the students were encouraged to express critical opinion about the program it was observed that they were equally critical of the course management issues such as the length of classes in the initial recordings (in the later recordings there are no complaints as either they got accustomed to the long sessions or started enjoying the program).

Students working at a regular pace and sincerely engaged: Observation through this method shows that the students are working at differing pace. There appears to be no compulsion to compete or finish earlier than others. Self access materials encourage work at students' own pace. This is supported by the varied- in terms of time- nature of the comments on each session. Within one session recording there were students working on unit 4 of CD1 and some working on unit 2 of CD2. In the first two sessions in the laboratory the majority of the students were observed advancing together; however, their pace of working on the program started varying soon after. Even in the second session some students were still finishing Unit One whereas a few others had finished it and gone on to attempt the exercises in unit 2 or 3.

Students understand the concerns of the evaluator: Automatically the topics which would be of special interest to a teacher evaluator are considered important by the students also. Appropriateness and content based comments are the most in number. These would be of greater value to an evaluator who is trying to match the contents of a program to the needs of the students rather than a materials designer undertaking a formative evaluation with the aim of improving the functionality of the program.

A very learner centred method: The method provided a lot of latitude for individuality. As an evaluation method it provides a whole range of individual likes and dislikes. Most detailed and individualized commentary is thrown up. However the subjective nature of the method does not preclude objectivity being observed by the students

in their comments. One aspect which would be of interest to the developer more than the teacher researcher was that not only were there fewer comments related to errors but also fewer elaborations or reasoning around usability related comments. Once they detected an error or faced technical or functional difficulty in the program, the students commented on it and moved on recognizing their own limitations and lack of expertise in programming such learning software despite some students being Computer Science majors. One conclusion is that perhaps this method does not capture usability issues as well as another method especially designed to do so such as an implementation log or a monitoring program that sends error reports, or for that matter a usability measuring questionnaire. On the other hand there may not have been many usability related issues because the program is a well designed one.

Greater awareness of learning: Although the method did not purport to teach students reflection or self awareness yet as a by-product of the use of these methods resulted in students becoming reflective of their performance. Section 5.3.1.1.4 illustrates this from excerpts from students' comments. The commentary is detailed and positive about students' performance. One example of how the students were closely monitoring their learning and making full use of the CD-ROM and applying it appropriately to their learning needs is evident in the reflective nature of the following comment:

98: Another thing I'd like to mention in referencing in Unit 4 is that it's often hard to understand the foreign names and the sources that teachers use like most of them I understood what the name was but I didn't understand the exact spelling of it and when you don't know the spelling you can't go on a website and put in the name or search for them on a search engine or go to the library website where you have to type in the author's name if you don't know the spelling then it's pretty useless knowing how the name is pronounced and instead of making the students guess the teachers should write the names down on the board and exercise 7 is quite useful for exactly the same reason because even if you don't know how to spell the name one should be aware of who or what the teacher is talking about and whether the teacher is referring to a book or a person or author for example [?] you should know that it's a book as well as a person you know even if you can't spell it you should know what it is so that later on when you do your research you should have some way to start off and okay at least we know it's a person we can go search for people... (WAV Session Five [Cd1]Paragraph 98 Node Appropriateness Student RP20F)

This excerpt starts on a negative note ending with the student rationalizing the importance of the learning content of the CD. It also provides evidence of some very basic yet practical learning point the student seems to have identified. There is also an embedded suggestion “...instead of making the students guess the teachers should write the names down on the board...” which is followed by the program in the next exercise. It is such comments that validate the use of this method as a means of mining deep and rich content-specific data.

ROP spontaneous in nature and RRWA reflective and tempered: It was observed that the Verbal accounts retained the immediate spontaneity of TAP protocol even though there was some time lag between the exercise and the recording. However there is no guarantee of their spontaneity because as soon as some distance occurs between the learner and the observer the observer loses control over when the recording is happening. The emails appeared more thoughtfully constructed. A greater discussion of this aspect follows later in section 5.4.2 where the variations of RP are being compared.

5.4.1 *The method's performance on the criteria of evaluation*

In this section the two variations of retrospective protocols are evaluated on the criteria established in chapter three of this thesis. Table 5.3 gives a graphic representation of ROP and RRWA and their performance on the evaluation criteria.

Rating Scale:
Negative: (-,-,-) very poor/ (-,-) poor/ (-) slightly poor
Positive: (+)ok/ (++) fair/+++good/++++very good

Table 5.3: ROP and RRWA Performance Matrix

Criteria	Retrospective Oral Protocols	Retrospective Written Accounts
1. Cost Effectiveness		
Time	Inexpensive (+++)	Inexpensive (++++)
Effort	Inexpensive they are just talking(+++) but training in think-aloud conventions required	Slightly more effort required on the part of the students to write the reflective account. (++)
Money	Inexpensive(++++)	Inexpensive (++++)
2. Ease of Use		
Preparation	Some preparation training needed	None required (++++)

	of both respondents and researchers(++)	
Implementation	Technical issues may occur (-)	Very easy(++++)
Data preparation	Transcription/coding (-, -,)	Only coding no transcription (++++)
3. Bias		
Researcher	No bias (++++)	No bias (++++)
Respondent	Yes respondent's Subjective opinion sought (++)	Yes respondent's Subjective opinion sought (++)
4. Ecological Validity		
Researcher Intrusiveness	Yes quite possible	Yes possible as the recipient of the mails.
Environment Intrusiveness		
5. Nature of Information		
Deep/Surface	Deep (++++)	Yes Deep personalised information (++++)
Objective/Subjective	Very subjective information (++)	Subjective nature initially but objectivity introduced through reflection (++)
6. Immediacy of Response		
Immediate/Delayed	Yes. Immediate (++++). General idea immediately available for practical purposes without having to be analysed.	Yes (++++). Results immediately available for practical purposes without having to be analysed
7. Usability Measure		
Satisfaction	Yes (++++)	YES (++++)
Effectiveness	Not so effectively (++)	YES they get time to reflect (++++)
Efficiency	No (++)	YES to some extent (++)
8. Robustness of Method		
Validity	Yes but too much variation 'waffle' (++)	Yes but since reflection has taken place, How much is actual and how much for effect (++)
Reliability	High (++++)	High (++++)
Productivity	Very High (++++). Longitudinal formative nature yielded a lot of points	High (++++). but mostly as a repetition of what was said in the oral think-aloud protocols but justification through reflection has been added.

The versions of retrospective protocols used in this study proved to be *easy and inexpensive* methods which required little effort to set up. Some familiarity with WAV functionality and sound equipment installation is necessary on part of the teacher evaluator, and a little effort is required to learn this if the evaluator is untrained in dealing with related hardware and software issues.

Respondent training in oral protocols and *think-aloud* is necessary to get good results from this method so the method is effort intensive in preparation for recording on WAV, but less so when emails are employed. Although some technical problems in the recording mechanism may occur occasionally this is not a sufficient

threat to drastically affect the rating for ease of use. For purposes of research data preparation is required but for the teacher evaluator just listening to the recordings and lifting relevant points will in many cases be enough. Emails need to be read and selected but no transcription is required, only coding for purposes of research.

Bias: The researcher has no control over what the respondents say in the protocols, so as far as raw data is concerned there is no researcher bias. As in all qualitative analysis, however, there is likely to be bias or subjectivity present at the interpretation stage. Respondent bias by the very nature of the method is high because this is a subjective method which is trying to elicit personal opinion, and will therefore take into account personal preferences and learning styles.

As far as the *ecological validity* of the method is concerned, there is no intrusion by the researcher and the researcher's presence does not affect the gathering of the data. However, the recordings and emails are made expressly for the teacher evaluator, and the student's idea of the listener/recipient may to a certain extent affect the message content. The artificiality of recording impressions into a machine influences the validity of the data to a small extent, as does the compulsion to record on WAV or write an email. This artificiality will also be present to more or less the same degree in responding to survey questionnaires. The fact that the same comments are repeated in data from other methods, however, suggests that ecological interference and researcher intrusiveness were relatively low in both versions of the method.

The *nature of information* from this method was highly subjective. Although respondents tried to objectify and rationalize their views the method also captured individual points of view, as it was meant to. Information was deep and abstract on occasion, when respondents tried to explain their viewpoints but at other times it could also be rather superficial and shallow. Certain comments were repetitious and quite meaningless and gave the impression that the subjects were commenting for the sake of it, because they were required to say something after each unit of work. The qualitative approach to data analysis of this method made it possible to identify the deep and shallow nature of different comments.

On the criterion of *immediacy of response* this method performs fairly well because an overall impression can be gathered by just listening to and reading the comments. No tabulation or extensive data preparation is required by the teacher evaluator. Depending on the number of participants in the study it can be a quick or long drawn

out affair. For the researcher data needs to be transcribed and analysed but an overall impression is possible from a first perusal.

As a *measure of usability* this method fared well on capturing *satisfaction*. The scale of *effectiveness* was better captured in emails which were summative (end of each working session) as opposed to the WAV recordings which were done formatively. In the emails the students get more time to reflect on their learning and their learning gains which they attribute to the *effectiveness* of the materials. Efficiency was not captured at all in the oral version because the students were recording while working on the materials. The *in-use* nature of these recordings resulted in the students not dwelling for long on problems of functionality, as they were eager to continue working. They do mention snags and programming flaws, but only in passing and without any elaboration. *Efficiency* was captured to a certain extent in the emails because the students commented in more detail on the topic.

This method is high on the scale of *validity* as it captures learner's *in-use* and *post-use* opinions. The validity of the data from this method is substantiated by the data from other methods. On the criterion of *reliability* this method will deliver the same or similar results if replicated in the same context. If the context differs then results will be different. On the scale of *productivity* this method scores highly as the nature of the method was longitudinal and formative therefore many opportunities at different intervals of time were presented to the students, resulting in detailed and rich commentaries.

5.4.2 *Comparison of the two variations of retrospective protocols*

A comparison of the two methods raises the question of whether the methods are complementing each other through validation or whether memory is playing a role. Comments match each other fairly well, meaning that the WAV recordings talk about the same issues as the emails from the same day, (although these tended to have greater grammatical accuracy and completeness of sentence structures). The emailed comments followed the aspects identified in the earlier verbal protocols and added more structure and fewer entirely fresh insights, although there is a greater aspect of reflection in the emails. On the scale of spontaneity and closeness to the experience the emails are post-use and a certain moderation of opinion and order born out of the passage of time has been imposed on them. The question arises as to how much the two methods may be impinging on each other. Is the validity of either method compromised if they are used in tandem? For instance if the student

has made a comment in the ROP and then mentions the same topic in the RRWA, does the use of the RRWA just reinforce earlier thinking or does it foster a fresh reflection on the topic? There was sometimes a marked similarity between the verbal protocol and the same session's email. On the other hand there was also evidence that further points were added in later observations, indicating that fresh ideas had come to the user retrospectively.

It was interesting to note that there were a few students who did not have very effective ROPs but who produced very good RRWAs (RP 13M; RP20F; RP10F; and RP11F). Girls appeared to be better at reflecting, and their emails were more thoughtful, and individual. The emails of two students had the same generic pattern and started with the same wording, with just a few new comments at the end suggesting that each had developed a template and was using it every time to write the email. Only two students were detected to be copying from their earlier work in this fashion (RP52M, who did so well meaningly, and RP 29M, who was probably playing the system). The teacher evaluator's personal knowledge of individual students, and the honesty with which they are likely to respond, can act as a benchmark to gauge the value/ validity of other students' comments. My own personal knowledge of the students leads me to believe that by and large they gave sincere views. I can make the assertion that shy speakers thrived in written reflection, revealing yet another positive aspect of RRWA.

5.5 Summary

In this chapter the data from two variations of retrospective protocols; ROP and RRWA, were analysed separately. After explaining the procedure followed for collecting data a detailed analysis and explanation of themes and codes is done first of oral protocols then of written emailed accounts. The main findings of the two methods are summarized according to the perspective of the evaluator. This is followed by assessing the method's performance on the criteria of effectiveness.

Having dealt with two qualitative methods, retrospective protocols in this chapter and focus group interviews in the previous one, I now turn to quantitative analysis and questionnaires in the next chapter.

CHAPTER SIX: QUESTIONNAIRES AS AN EVALUATION METHOD

6.1 Introduction

This chapter will discuss the use of two kinds of evaluation questionnaires: PLUM (Programme on Learner Use of Media) and SUMI (Software Usability Measurement Inventory). The first section discusses the results of the PLUM pre program and post program use questionnaire. The second section will introduce the methods of analysis used for SUMI and the results of using it. In the fourth section a summary of findings from the two questionnaires which would be of interest to a teacher evaluator are presented, and then both questionnaires are evaluated according to the core criteria.

6.2 PLUM questionnaire

The two PLUM questionnaires, pre and post program use, can be found in Appendix 6.1 and Appendix 6.2. Chapter Three section 3.6.3.1 explains the procedure of administering the method.

6.2.1 Pre-use questionnaire

The PLUM pre-program use questionnaire has only one open ended question:

What are you hoping to learn from doing this exercise? Please note down the knowledge or skills you would like to improve in the course of this session.

The respondents' answers to this question recorded their expectations from the program. These *Expectations* were coded into three categories or main points, namely: *General*, *Academic* and *Professional*. The number of points in each category is given below:

Expected General Gains: 16

Expected Professional Gains: 9

Expected Academic Gains: 23

This pre use questionnaire was given back to students at the end of the program along with the post use questionnaire which inquired whether their expectations had been met.

6.2.1.1 General

These comments were mostly about the global general benefits that the students expected to gain from working on the EASE program during their communication skills course. Personality gains, confidence building, and general improvement in communication for everyday purposes were all subsumed under this category of expected general gains as exemplified by the following:

22: I would like to polish my personality by practising writing and doing speaking sessions. Q-038

54: Not only will this allow me to successfully discern what others are saying but shall help in presenting me as a confident and self assured individual. Q-020

175: I also want to be a good social speaker and be able to speak confidently. Q-016

79...Listening is the first part of an effective communication. I hope to improve my listening abilities by using and implementing this software. Furthermore, I hope that in that process I can expand my knowledge. Q- 040

217: First of all by doing this exercise we can improve on our listening and we can listen effectively. This will basically improve our speech as well and we will gain confidence. Speaking in public is the main point. I feel shy in public speaking. I think I will improve on this after I finish this course. I will improve my accent and can present in front of people as a effective speaker. Q- 050

6.2.1.2 Professional

Professional and career benefits were also a recurring theme. A few examples from the comments follow.

22: I hope to learn the different forms of communication used in the professional world. Q-038

79: The art of communication plays a vital role in virtually all aspects of our lives. Be it conversing with friends or be it presenting point of view in a business meeting, there is no denying the fact that effective communication skills can make the difference between success and failure. Q 040

6.2.1.3 Academic

The majority of the student expectations from the program pertained to academic improvement.

264: I am hoping to improve my analytical and organizational skills in this course. That would enhance my listening, reasoning, speaking, evaluation and self criticism skills. Q- 039

54: What daunts me most is when a long lecture is followed by a series of questions that need pinpoint accurate answers. Hence my aim through the course will be to build a solid bridge between my memory, mind and mouth, such that communication is easy. Q-020

60: It will also give me a chance to master the art of note taking during a lecture. Another expectation is that my vocabulary should improve. Q- 022

90: I am attending this exercise in a hope to increase my listening power that will help me in my future lectures of other courses. Moreover this course will help to improve my presentation skills. It will help me to enhance my confidence in speaking. Q- 034

232: With the help of this course I am able to take notes of lectures without leaving out the important points, to be able to improve listening skills and understand foreign accents, to improve concentration during a lecture and to improve vocabulary. Q- 029

These comments will be matched to the first question on the post use Questionnaire.

6.2.2 PLUM post use questionnaire

The PLUM post use has four questions. Questions 1, 3 and 4 have open ended sub questions and responses were analysed qualitatively. As the answers to these open ended questions were varied only those points which were made more than once have been included for discussion. Question 2 was analysed quantitatively using Microsoft Excel and factor analysis using SPSS. (Cf: Chapter 3 Section 3.6.3.1.1).

6.2.3 Analysis of Listening to Lectures

The post-use questionnaire was administered separately for the two CD-ROMs as the student impressions of the two CD-ROMs were quite different. The findings below concern EASE volume 1: *Listening to Lectures*.

6.2.3.1 Qualitative analysis

6.2.3.1.1 Question 1:

Look back at what you wrote for the Pre program Question and note down: (a) what you learned that you hoped to learn: (b) what you did not learn that you hoped for, and (c) anything you learned that was unexpected

The answers to the first sub question 1a were mostly positive, to 1b and 1c were very varied as was expected because there were probably as many different expectations as there were respondents. The points that emerged from the data and the number of times these were mentioned across the respondents are given below. The points that were mentioned at least twice have been included in the discussion. Most comments concerned improvement in listening or communication skills.

Question 1 (a) *what you learned that you hoped to learn:*

In response to this question students commented on improvement in communication skills and listening to lectures(20), an improvement in understanding lectures(14), note taking abilities(11) and improvement in concentration span(3). Students also mentioned recognizing foreign accents, and improvement in vocabulary as areas of improvement.

1a: Proper note taking, improved concentration and ability to listen and recognize foreign accent. (Para 356 Q-029)

1a: My vocabulary and note taking skills have certainly improved. (Para 571 Q-041)

A sample of other comments follows:

1a: to improve my note taking and understanding of lectures. (Para 8- Q-001)

1a: I hoped that I would improve my lecture listening skills and in short it did. (Paragraph 64-7 Q-006)

1a: I expected that my ability of grasping what the lecture is trying to convey would be enhanced and through practice and a lot of listening I believe the purpose has been served. (Para 115 Q-011)

Question 1 (b): *what you did not learn that you hoped for:*

A variety of points emerged from the data in response to this question. Out of 38 questionnaires analysed some comments mentioned felt that the CD-ROM failed to build confidence (6), others said that it failed to teach effective communication (5) and that it failed to teach them how to interact(5). They also said that it did not teach class participation (2) or persuasive writing (2) nor help in making better notes (2).The answers to

Question 1b seem to be based on the students' expectations recorded in the pre-use questionnaire and not the actual use of the program. Some interesting comments are given below.

1b: the program actually taught a lot more than I had expected. Q-001 Para 9

1b: I expected the first CD to guide us on class participation and how to make your presence felt in the class but didn't come anything as such. Q-011 Para 116

Question 1(c): *anything you learned that was unexpected:*

Out of 38 respondents, 9 did not respond to this question. The most commonly recurring point was that note taking was an unexpected outcome (7), followed by understanding of the structure of lectures (5) and familiarization with foreign dialects and accents (5). Other comments pertained to improvement in vocabulary (4) knowledge gained about sciences and socio-political subjects (4) and awareness of grammar.

1c: that the expressions and body language say a lot more about people than words. Para 10 Q-001

1c: Taking better notes. (Paragraph 67 Q-006)

1c: The structuring of the lectures was something which I didn't expect, but eventually came across it. Para 117 Q-011)

The above analysis of question one and its sub parts corresponds to the question asked in the pre-use questionnaire and establishes that subjects' expectations were fulfilled in relation to effective listening and the development of note taking skills. Understanding lectures was the predominant positive answer regarding what was learnt, while negative answers were mostly concerned with the subjects' lack of *confidence*, and the lack of provision for personality development and interaction.

6.2.3.1.2 Question 3:

This program is meant to help you improve your academic listening skills. Could you please comment on any improvements you are aware of in: a) Your knowledge of how lectures are structured: b) Your note taking skills: c) Your approach to listening to lectures: d) ways in which the program could have done more to help you?

As the third question asked very specific questions relating to the learning effectiveness of *Listening to Lectures*, the answers are also specific. 38 questionnaires were analysed, but some responses were missing for each sub-question (3a=36, 3b=36, 3c=34 and 3d=30).

Question 3a: *Could you please comment on any improvements you are aware of in: a) your knowledge of how lectures are structured:*

22 students felt that the program improved their understanding of lecture organization whereas 13 identified different techniques used by lecturers. 7 students felt that their understanding of lectures improved because of this program. One unique comment was about the length of the lecture clips that were used.

3a: There should be a large lecture then the question answer session. Q-026 paragraph 314

Question 3b: *Could you please comment on any improvements you are aware of in: b) Your note taking skills:*

Some students felt that the quality of their notes improved (16) because they were using abbreviations (4) and their improved understanding (6) and relevant note making (11). Some remarked on the speed of their note making (4) and improved retaining power (2) because of effective note making.

Question 3c: *Could you please comment on any improvements you are aware of in: c) Your approach to listening to lectures:*

Most students reported that they were able to discern the key points (10) in the lecture through greater awareness of structure, organization, tone and emphasis(13). They also felt that repeated listening practice had improved their concentration spans (6). A few, however, felt that little or no improvement had taken place (3). One student noted increased awareness of lecturers' uses of humour.

3c: I am now able to judge where the lecturer is being serious and therefore where to pay more attention Q-014

Para 152

Question 3d: *Ways in which the program could have done more to help you?*

This question yielded varied answers but the greatest degree of consensus concerned the need for a greater variety of exercise types (5), a testing scheme to assess performance and to make the program challenging (5),

and more interaction and opportunities to speak (10). Other comments asked for more note taking exercises, fuller videos and more frequent use of the dictionary function (6). One comment which illustrates the competitive nature of participating students is given below:

3d: If it had a marking scheme to calculate our grade after answering the exercise .Q- 022 Para 269.

This concludes the discussion on Question 3. It is important to note the similarities of these points to those made in other methods.

6.2.3.1.3 Question 4

Question 4: *Would you want to use it again? Please say why, or why not:*

The answers to this question suggest that of the 38 respondents, two thirds felt that they had learnt all that the CD-ROM had to offer and would not like to repeat.(19). However there were others who were prepared to do it again to improve English (7). Some students were categorically one sided in their opinion whereas a few said that they would use the CD-ROM again “if” less time could be spend on it or “if” new clips, more challenging exercises etc., were added. Therefore these “yes and no” answers are cross listed in both categories.

4: No! The program certainly was useful but I am fonder of listening to teachers. We can challenge them and exchange opinions. Q- 001 Para 17

4: No, the program is more of a waste of time. It is like the 6th grade comprehension and grammar (vocabulary) exercises. It might have improved our listening skills to some extent but not our communication and talking skills. What good does it do to you if you are a good listener but cannot talk or converse impressively with another person? Para 228 Q- 018

4: No, because it is just like reading a book. Now I know for once and for all that what is in there. Q- 027

4: No. I think going through each unit once is enough. The CD is easy to operate and it is any easy way to learn. Using the CD again would mean memorizing the answers to each question. Q-029 Para 365

This concludes the qualitative section of the PLUM post-use questionnaire for *Listening to Lectures*.

6.2.3.2 Quantitative analysis of PLUM post use questionnaire

Question two asks respondents to evaluate the CD-ROM on an intensity scale of 1-5 where 1 is strongly agree; 2 is agree; 3 is neutral; 4 is disagree; 5 is strongly disagree. It has seven sub questions:

Question 2: To what extent do you agree with the following descriptions of the program?

(2a): Easy to operate (2b) Enjoyable to use (2c) Provides good support for the exercise (2d) Provides good advice on how to approach the task (2e) Helps you learn (2f) Fits well with the rest of the course (2g) Well worth the time spent on it.

This question's component parts were first analysed using Microsoft Excel. The pivot chart function of Excel was used to create the graphs. I have selected tables and graphs of four sub-questions to present here.

Table 2a: Easy to operate

Strongly Agree	22
Agree	14
Neutral	1
Strongly Disagree	1
Grand Total	38

The table reveals that majority of students felt that the *usability and operation* of the program was easy. This is the limited measure of usability that this questionnaire offers.

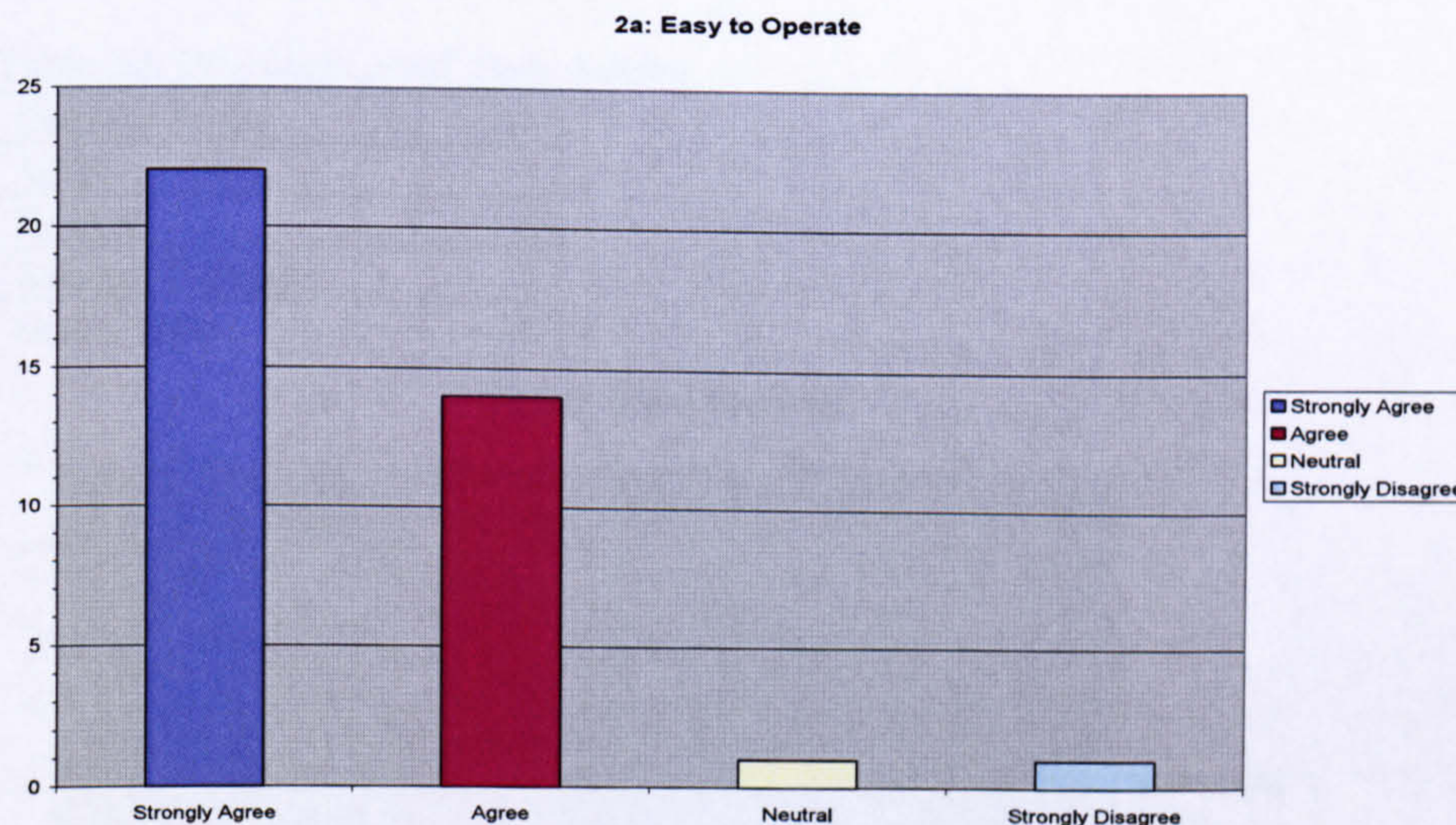
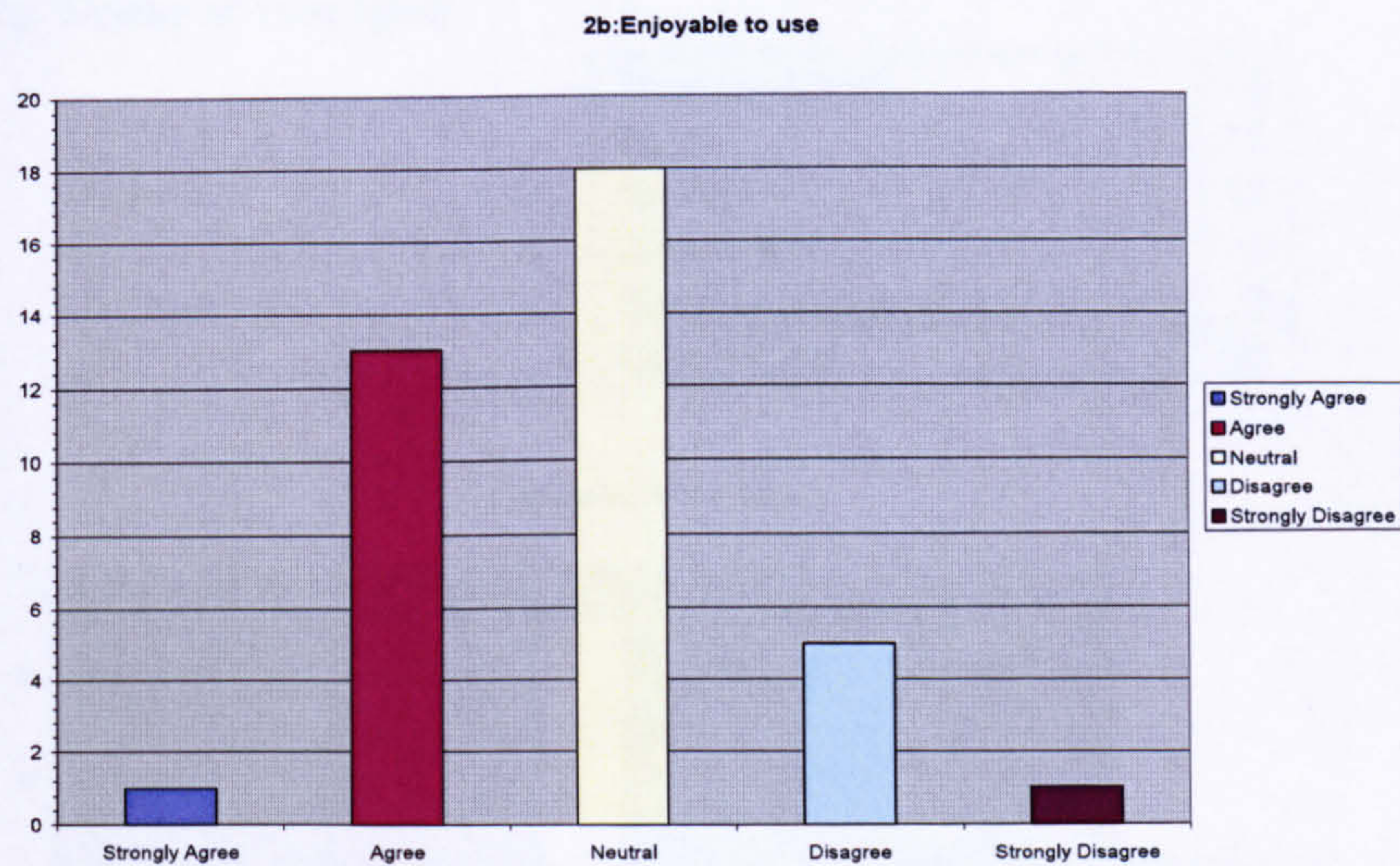


Table 2b: Enjoyable to use

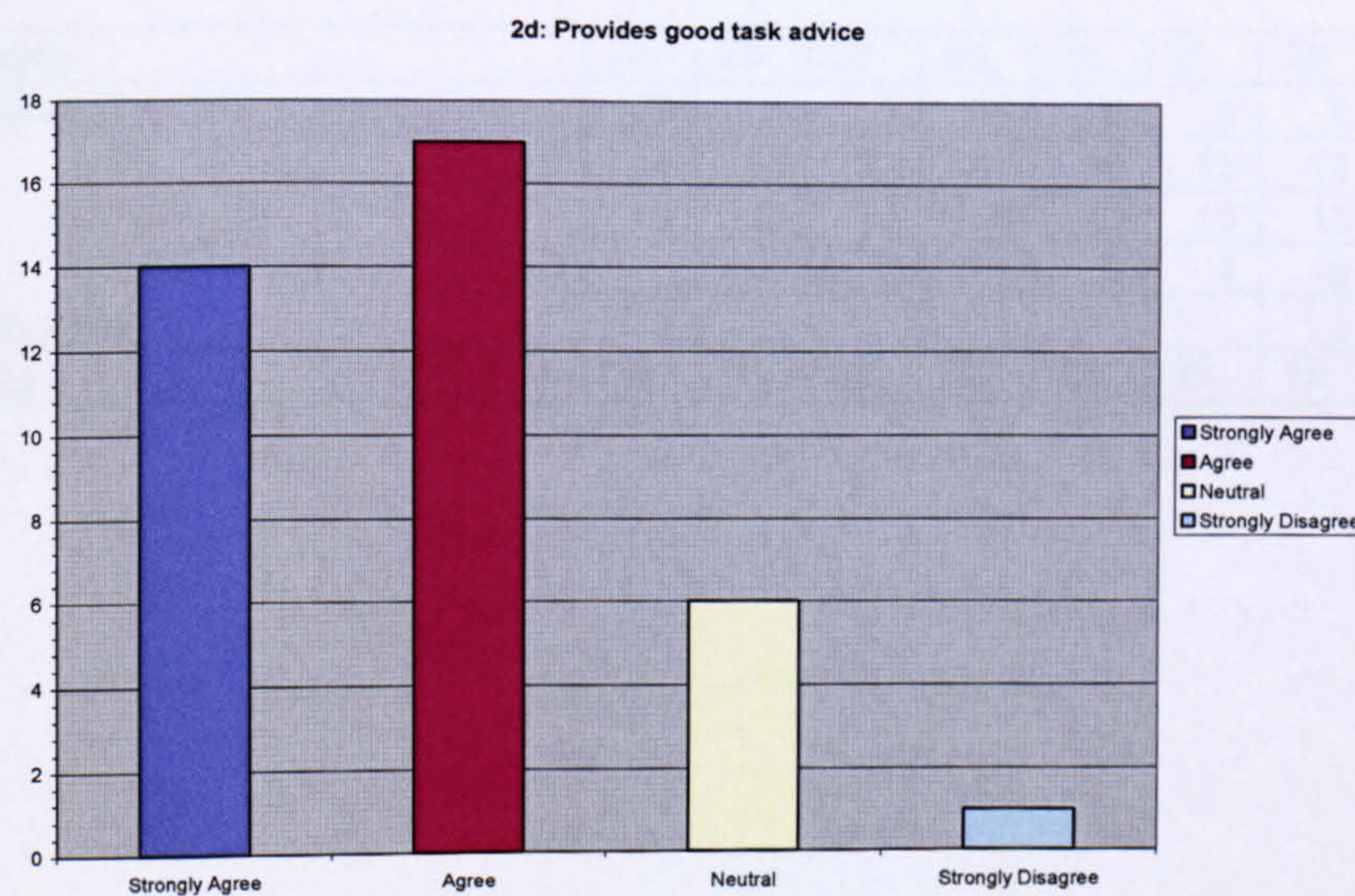
Strongly Agree	1
Agree	13
Neutral	18
Disagree	5
Strongly Disagree	1
Grand Total	38



In answer to Q 2c which asked whether adequate support was provided for the exercise there is positive agreement with only ten respondents staying neutral. The respondents were also in strong agreement that the program offered good task advice Q2d.

Table 2d: Provides Good Task Advice

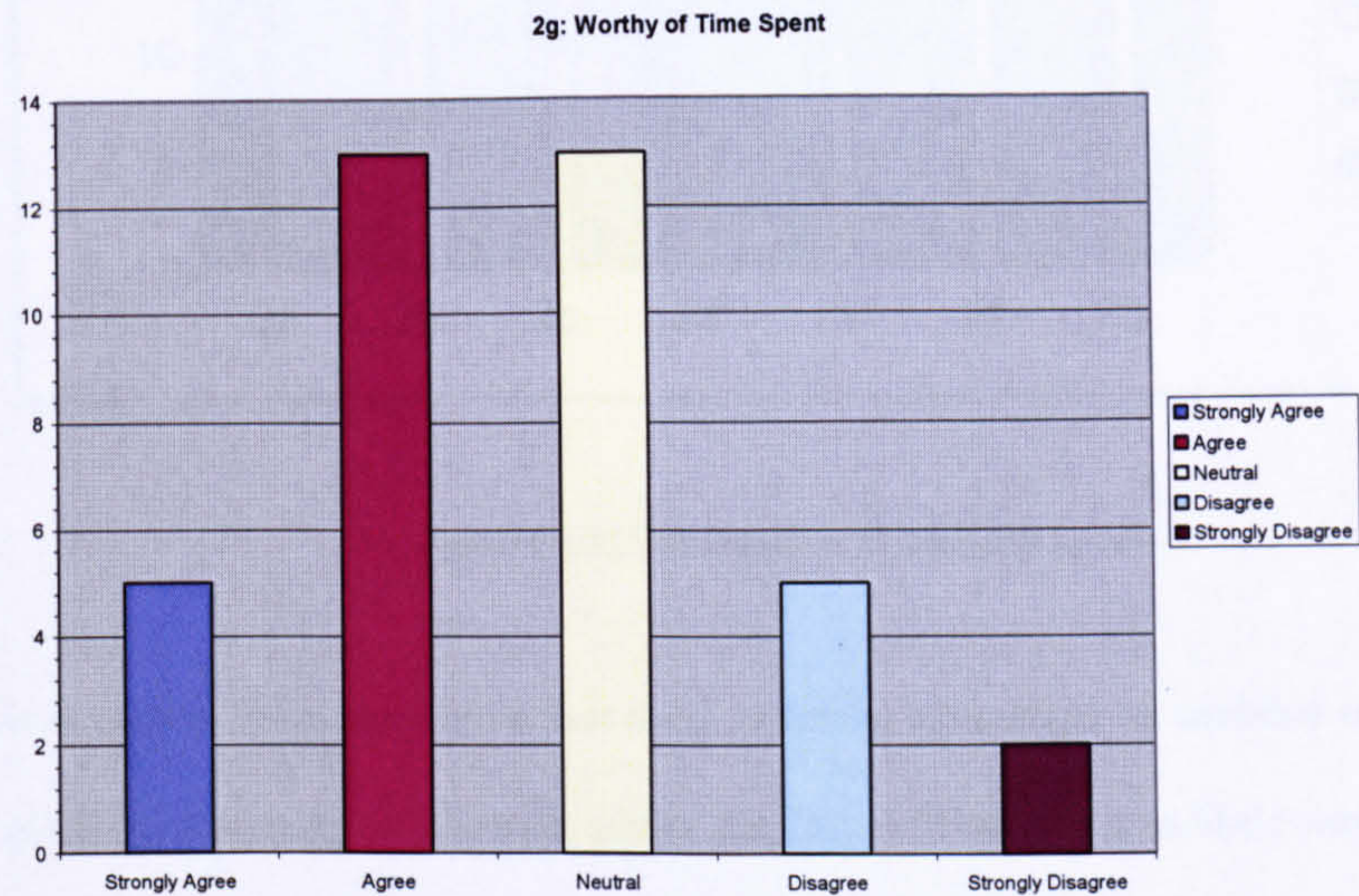
Strongly Agree	14
Agree	17
Neutral	6
Strongly Disagree	1
Grand Total	38



The respondents verdict is mixed on whether it was worthy of the time spent.

2g: Worthy of Time Spent

Strongly Agree	5
Agree	13
Neutral	13
Disagree	5
Strongly Disagree	2
Grand Total	38



Summary of Findings: Collating the above data gives the following result:

Table 6.1 PLUM Question 2 *Listening to Lectures*

Opinion Scale	2a	2b	2c	2d	2e	2f	2g
Strongly Agree	22	1	2	14	6	8	5
Agree	14	13	24	17	19	13	13
Neutral	1	18	10	6	10	13	13
Disagree		5	2	1	3	4	5
Strongly Disagree	1	1					2
Grand Total	38	38	38	38	38	38	38

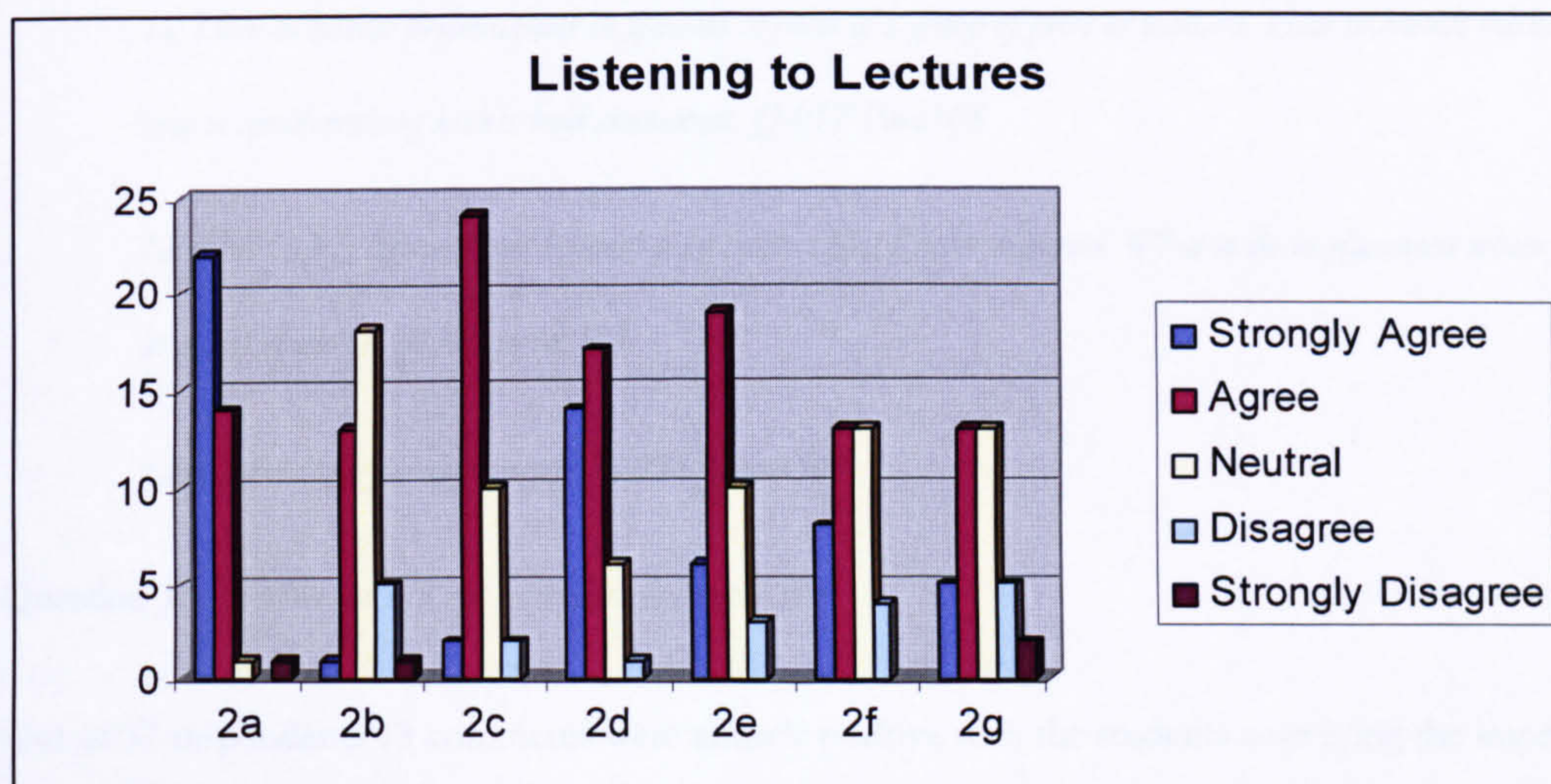


Figure 6.1 PLUM Question 2 *Listening to Lectures*

It is evident from the above that there is strong agreement on usability of the program (2a) and agreement on good exercise support and task advice (2c, 2d). Affective factors like course appropriateness (2f) time well spent (2g) and enjoyable to use (2b) received relatively less approval.

6.2.4 *Analysis of Seminar Skills*

The second CD-ROM was evaluated using the same post use questionnaire and yielded the following results:

6.2.4.1 *Qualitative analysis*

6.2.4.1.1 *Question : same as in 6.2.3.1*

Each of the sub questions yielded different responses which have been coded on the basis of the points made.

The frequency of each point is counted and given in the sections below.

Question1a: *a) what you learned that you hoped to learn:*

The answers to this question show that out of 37 respondents 30 felt that their objective of improving their presentation skills had been met. Other points concerned improved use of audio visual aids (4), improved speaking ability (5) improved understanding of the purpose of seminars and the way they are conducted (5), greater awareness of the needs of audiences (2) and gains in confidence (2) being able to distinguish between seminars and lectures. A few examples of comments are given below:

1a: How to handle presentations in speeches in front of a group of peers or teachers. How to handle mishaps and how to avoid making useless bold statements. Q-017 Para168

1a: This CD is different and I learnt a lot from it that I never expected. What to do in situations where you are at a loss of words. Q-029 para 333

1a: Learned everything, honestly! Q-038 para 454

Question 1b: *b) what you did not learn that you hoped for:*

Out of 37 respondents 13 comments were entirely positive with the students conveying the impression that they learned more than they hoped for. A fairly frequent criticism (repeated 5 times) was that their expectation of speaking practice was not fulfilled. While another set of comments (6) were about how they hoped to learn enhancement of presentation content and delivery and amalgamation of source materials. A few of the comments from which these points were derived are given below:

1b: Not much practical work was done so not really sure how speaking in front of a group will be easier. Q-005 para 52

1b: This program did not provide sufficient practice of spoken skills. Q016 para153

1b: Did not find ways to improve my confidence and to become a persuasive speaker.

Question 1c: *(c) anything you learned that was unexpected*

The student comments pertaining to presentations (9) showed that they developed an awareness of common mistakes students make in presentations; learnt definitions of seminars and presentations and to distinguish between them; learnt new techniques and a lot of useful tips; learnt how to overcome embarrassing situations in a presentation; learnt how to calm nerves before a presentation and to organize notes beforehand. They also learnt the use of audio-visual aids in different situations (4), about effective introduction of seminars and their organisation (2), and learnt to recognise different accents. Some comments are given below:

1c: How to use media in presentation and how to calm your nerves before a presentation. Q-018

1c: How and why to get nervous. Q-021

There is the possibility that other students are not filling out the questionnaire very seriously for there are instances where some sub questions are left blank.

6.2.4.1.2 Question 3:

This program is meant to help you improve your academic presentation skills. Could you please comment on any improvements you are aware of in: a) Your knowledge of how presentations are structured: b) Your speech making skills: c) Your approach to audio visual aids used in a presentation: d) ways in which the program could have done more to help you?

Question 3a: *Could you please comment on any improvements you are aware of in: a) your knowledge of how presentations are structured:*

In answer to this question the students felt that they had achieved greater awareness of structure and organization (25) and proper use of audio visual aids (7) and greater confidence in speech making (6). A relatively small number felt there was no or little improvement (3) and one student commented that it was not possible for the program to give actual practice in making presentations. A positive comment is given below which is unique in comprehensive content:

3a: Now I know the basic ingredients of a good presentation which are content, preparation and confidence. Q-040

Question 3b: *Could you please comment on any improvements you are aware of in: b) your speech making skills:*

The analysis of answers revealed overall agreement that there was improvement (32) in the delivery aspects such as appropriate style, level of formality, tone and eye contact (10). There was improvement due to increased confidence (10) which could be attributed to awareness of preparation and preparing backups (4) and better speech making abilities due to recording and hearing their own voices (8). There were some negative points (12) made by students which said that either there was no improvement (6) or it was difficult to assess improvement as there was inadequate speaking opportunity provided by the software (6).

Cryptic comments like the following show the ambivalence felt by the students about whether there was any improvement in their speech making.

3b: Not much but yes to some extent. Q-031

3b: I tried to improve and I think I did well in my exercises about voice recordings. Q-032

3b: I became familiar with the language which I must use when giving a presentation and how I should tackle the problem that I feel confident and be fluent in language. Q-034

Question 3c: *Could you please comment on any improvements you are aware of in: c) Your approach to audio visual aids used in a presentation:*

The points made in answer to this question were mainly positive. Students commented on how they learnt the importance of using audio-visual aids to create impact (17). They learnt effective structuring of audio visuals OHT slides (13) learnt to use different audio visual aids for different situations (7). Some representative comments are given below:

3c: The number of ways that visual aids can turn a presentation is enormous but we should not become over reliant on them. Q-002

3c: This has been one of the plus points of the software, the way they explained the use of audio visual aids in presentation. Q-040

Question 3d: *Ways in which the program could have done more to help you?*

In answer to this question the most frequently made points were about more hands-on workshop style practice of both making presentations (11) and public speaking exercises(6). Having more video content of complete presentations was also a repeated suggestion (7). A small number also felt that there should have been more examples of how to work with visual aids(2) and another two points were made regarding improving the interface and layout of the CD. However a significant number (9) felt that the CD-ROM was fine as it is and couldn't do more to help them.

A few sample comments are given below:

3d: By asking us to organize more data as to a real presentation. Q-027

3d: It should have answered the question; how to begin and organise your research? Q-041

6.2.4.1.3 Question 4

Question4: Would you want to use it again? Please say why, or why not:

The majority (24 out of 37) felt that they would like to use the program again thus giving a clear verdict on the program and its perceived efficacy. Whereas some students (7) felt that using the materials once was enough and nothing will be achieved by repeating it.

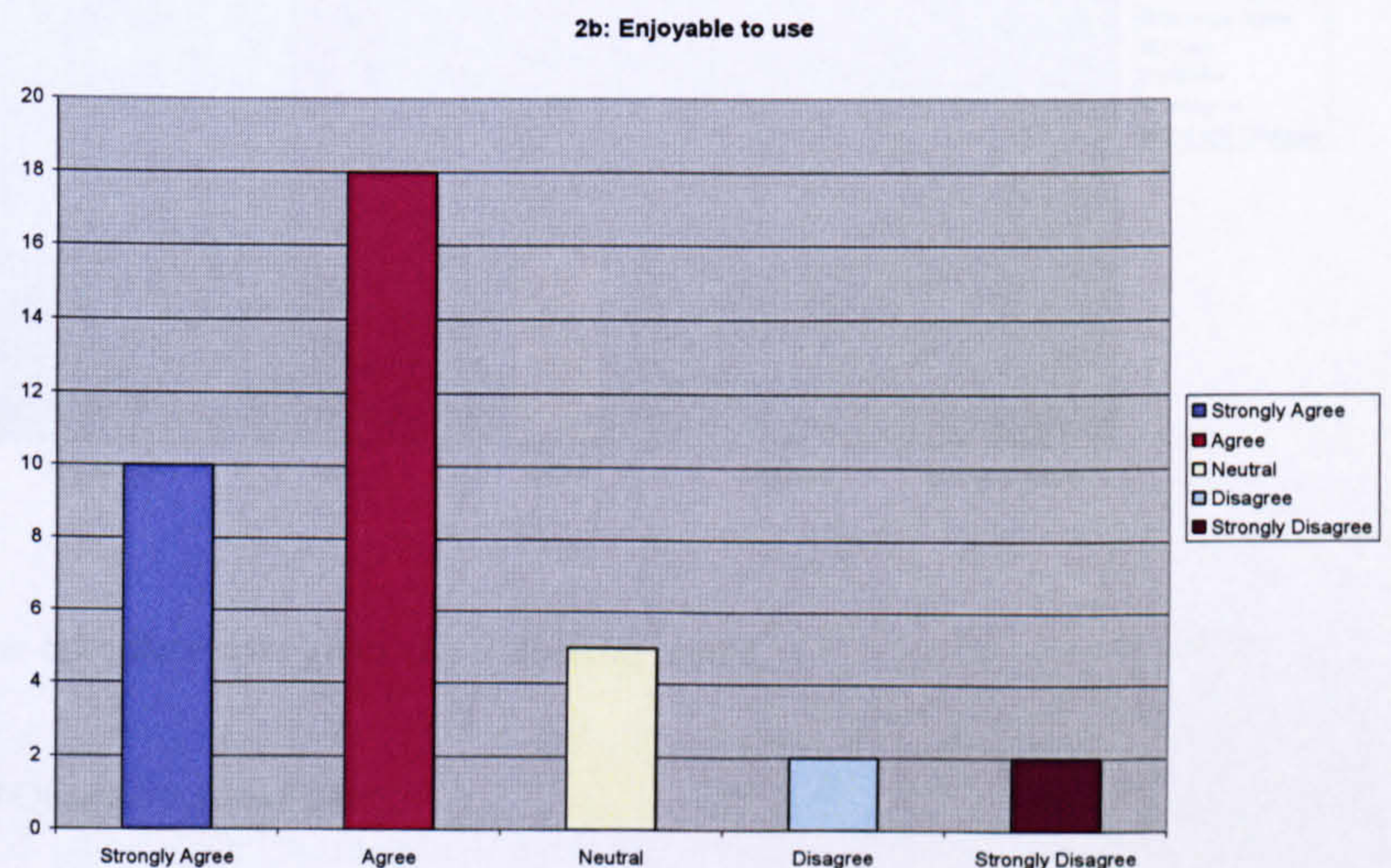
4: I wouldn't like to use the exact same program again as I feel that I have learnt everything that I could have from it. If I am given a similar programme then I may like to try that. Q-016

4: Yes I would but a more advanced level. I don't see any point in repeating the same CD. Q-023

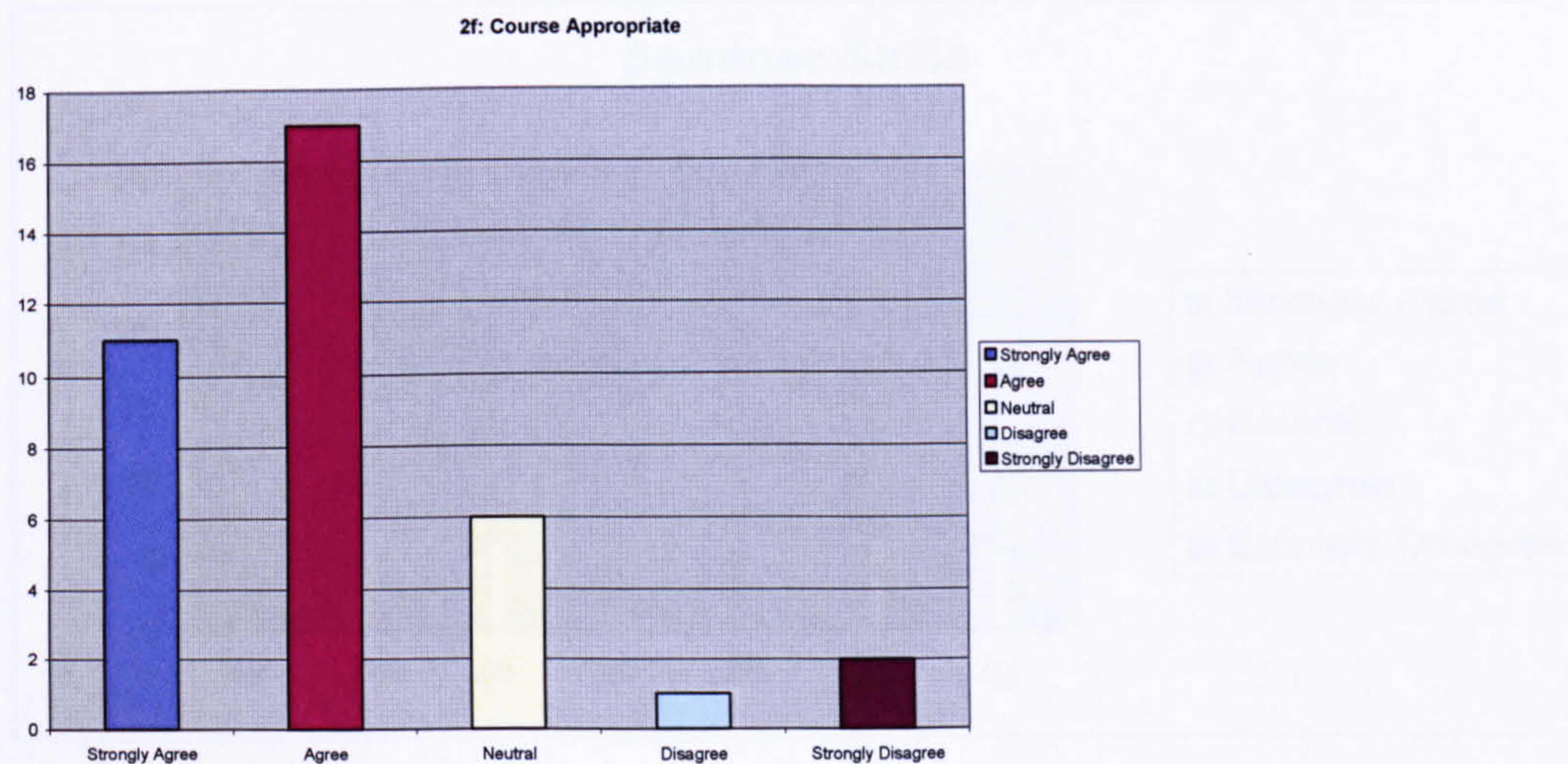
6.2.4.2 Quantitative analysis of PLUM post-use questionnaire Seminar Skills

Answers to question 2 were more positive than for the first CD. The students clearly stated that this was an enjoyable learning experience and was well worth the time spent.

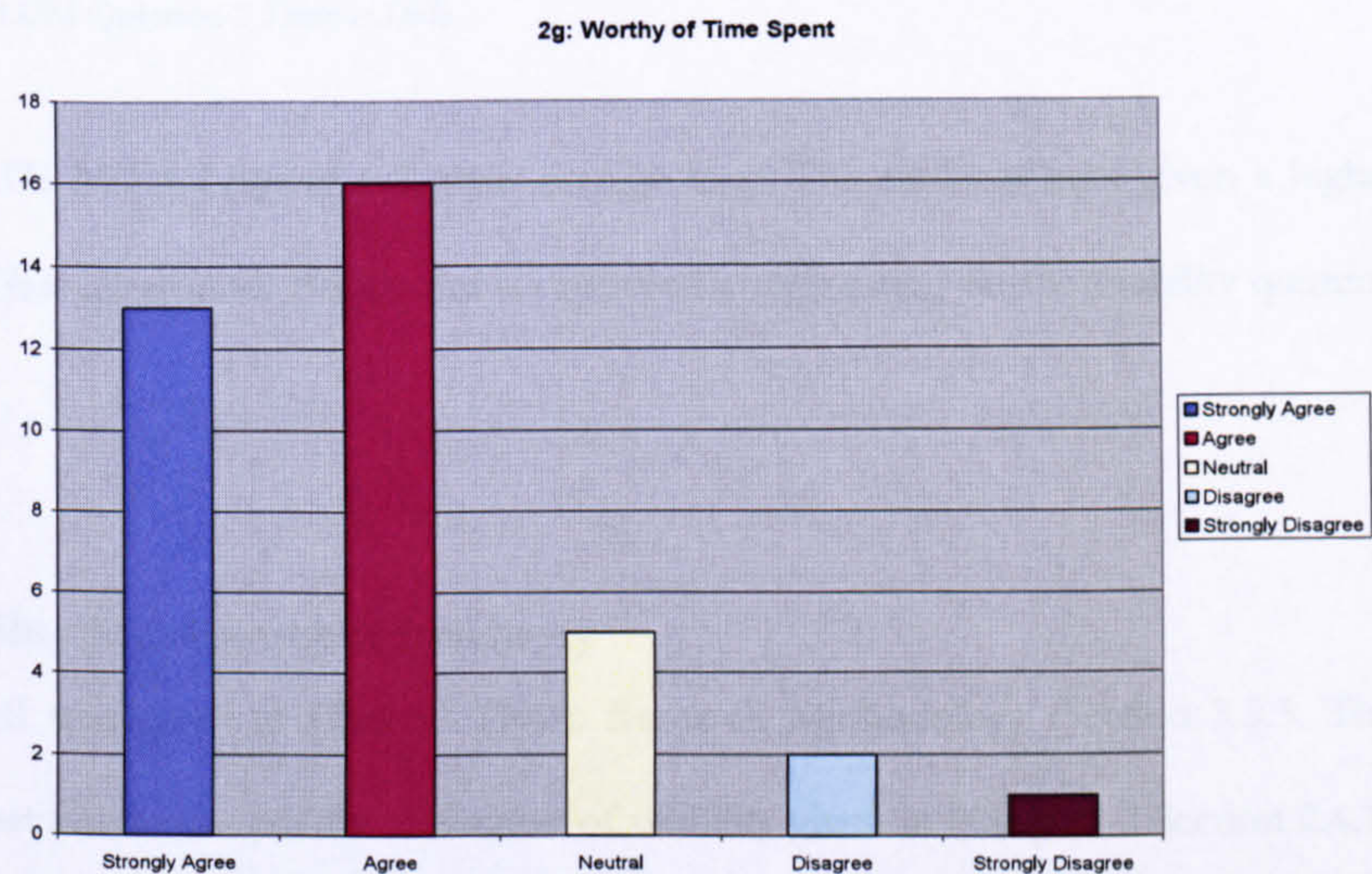
2b: Enjoyable to use	
Strongly Agree	10
Agree	18
Neutral	5
Disagree	2
Strongly Disagree	2
Grand Total	37



2f: Course Appropriate	
Strongly Agree	11
Agree	17
Neutral	6
Disagree	1
Strongly Disagree	2
Grand Total	37



2g: Worthy of time Spent	
Strongly Agree	13
Agree	16
Neutral	5
Disagree	2
Strongly Disagree	1
Grand Total	37



Summary of Findings: Collating the complete data gives the following result.

Table 6.2 PLUM Question 2 *Seminar Skills*

Opinion Scale	2a	2b	2c	2d	2e	2f	2g
Strongly Agree	27	10	11	13	11	11	13
Agree	8	18	19	21	23	17	16
Neutral	1	5	6	1		6	5
Disagree		2	1	1	2	1	2
Strongly Disagree	1	2		1	1	2	1
Grand Total	37	37	37	37	37	37	37

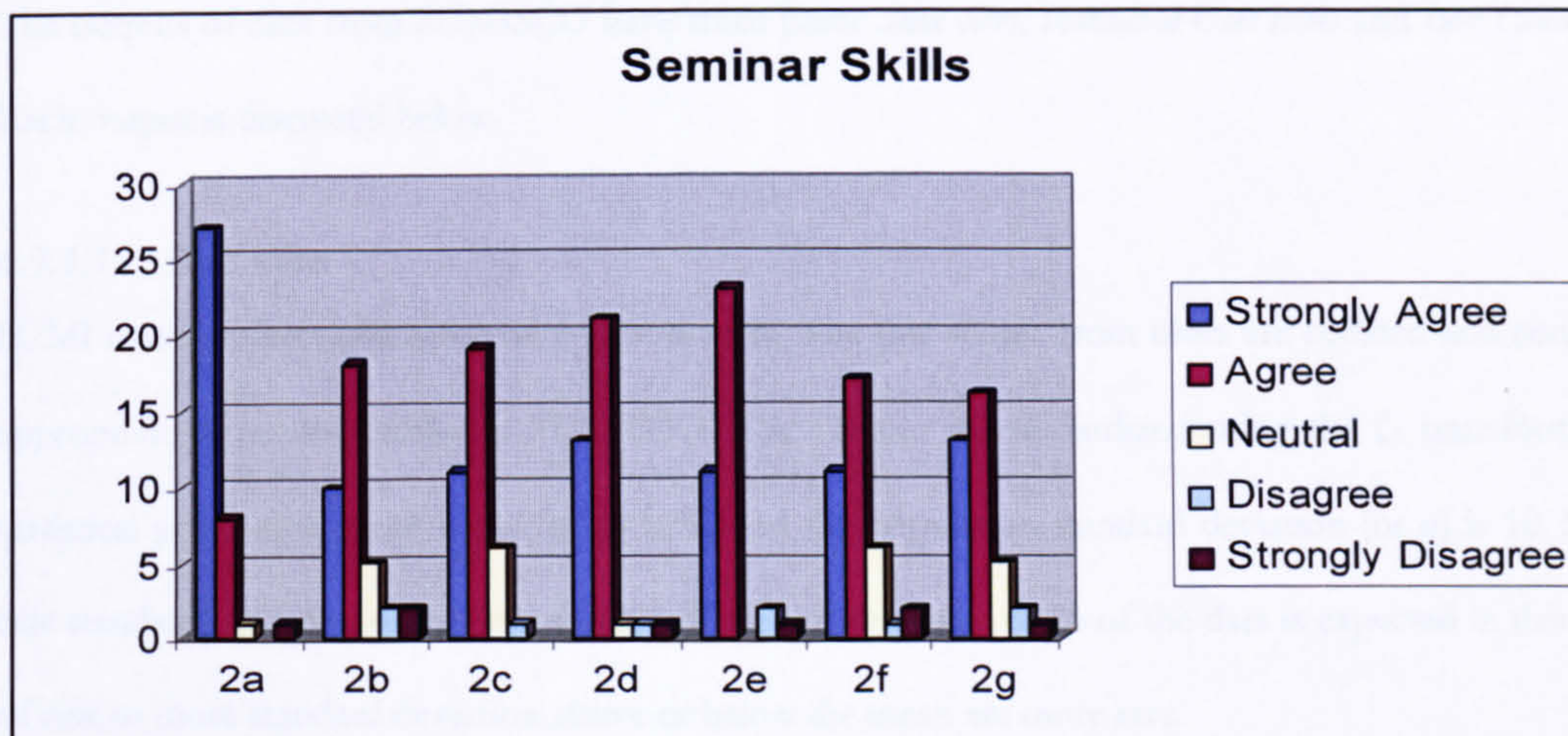


Figure 6.2 PLUM Question 2 *Seminar Skills*

Significant differences from *EASE Vol: 1 Listening to Lectures* emerge here. The students have given a higher rating to *EASE Vol: 2 Seminar Skills* all around. Both volumes received a high rating on the usability question (2a).

6.3 SUMI: Software Usability Measurement Inventory

A complete description of SUMI was given in Chapter Three: Research Methodology (Section 3.5.3). This inventory aims to measure user satisfaction as per the definition of usability given in (Chapter 2 Section 2.4.1). User satisfaction has been subdivided by the developers into five aspects namely: *Efficiency*; *Affect*; *Helpfulness*; *Control*; *Learnability*. (Cf: Chapter 3 section 3.7.5.2.1). These are measured by the five SUMI scales. There is a sixth SUMI scale which is 'Global usability' which is a general satisfaction measure.

6.3.1 Interpreting SUMI

SUMI can be analysed and interpreted using the scoring program SUMISCO which has been custom designed for SUMI by HFRG³. The following information is based on the SUMI User Handbook 2nd Edition (Kirakowski 1998:29-40).

³ Human Factors Research Group, University College Cork, Ireland

The outputs of data from SUMISCO have three parts: *Scale cores; Individual User scores* and *Item Consensual analysis*. Each output is discussed below.

6.3.1.1 *Scale scores*

SUMI contains five subscales and a Global scale. The raw scores from users are collated and compared to the appropriate normative tables in SUMISCO. The Output is standardized using the ζ - transform so that the statistical population mean score (or μ) is 50 and the population standard deviation (or σ) is 10. Scores within one standard deviation of the mean ($\mu \pm 1 \sigma$) are not very rare: 68 % of the data is expected in this band. Scores of one or more standard deviation above or below the mean are more rare.

Two statistics that are used by SUMISCO are measures of central tendency of the sample. The first one is computing the median which is more 'resistant' to 'outlying' observations in the data as compared to computation of the 'mean' which is 'unresistant' to 'outlying' observations. The second kind of statistical measure used by SUMISCO is dispersion measures. SUMISCO outputs two kinds of dispersion measures: 95% confidence intervals, and the 'fences'. 95% confidence intervals (UcL and LcL) where the true mean or median is likely to be in the sample 95% of the time if the study is replicated repeatedly. According to the SUMI handbook if the data is collected from a homogenous group the difference between 95% confidence intervals should be less than 10 for each scale (Kirakowski 1998).

Fences are denoted by UF and LF in table 1 of the SUMISCO Output report (cf: section 6.4.2/3). Fences are built around the median and are virtual numbers that depict the numerical limits between which 95% of the data can be found if the data distribution is 'normal' (Kirakowski 1998). Fences can be used to judge the normality of the distribution. Kirakowski (1998) suggests that there are two scenarios which need to be considered: 1) cases where some data exists well beyond a fence; 2) cases where the fence is much beyond the highest or lowest score on the scale. In situation 1) there is the likelihood of data being spread out with one or more outlying values (technical term; platykurtic). In situation 2) the data can be excessively grouped around the mean or median (leptokurtic). "If a fence on one side of the distribution is well beyond the final score on that side, it may also mean that the users are unevenly distributed around the median: this situation is found in datasets where the median and mean are in substantial agreement"(Kirakowski 1998:32).

6.3.1.1.1 *Interpreting the scales*

The five subscales of SUMI comprise 10 different questionnaire items. Each item is weighted by the contribution it makes to the subscale it belongs to. The five subscales are *Efficiency*; *Affect*; *Helpfulness*; *Control*; *Learnability* and the global scale *Global*. If there are some scores that are below 50 in the subscale then the software is poor in usability in that aspect. Subscales at or below 40 may require redesign. Good software will achieve a score of 60 or above (Kirakowski 1998). An explanation by Kirakowski (1998) of what high or low scores can mean is given next.

Efficiency: How efficient the program support is that enables the user to get their work done: *Low Efficiency:* Users do not always know what to do next, the software is slow and sometimes it works in a strange, inconsistent way. *High Efficiency:* The software supports and helps users in their work, it works at the users' pace in a predictable, consistent manner.

Affect: This scale measures how likeable the product is (stress-free usage): *Low Affect:* Users do not like this software, it is awkward to use and it makes users feel tense and frustrated. *High Affect:* Users enjoy their sessions and find the software mentally stimulating to use, satisfying and attractive.

Helpfulness: This scale measures the degree of information about the product in the product itself. *Low Helpfulness:* The messages and helping functions of the software are not very helpful, no information is consistently given and users are not helped to recover from their mistakes. *High Helpfulness:* This software communicates clearly, users can understand the way it works, and it gives them helpful hints and instructions.

Control: This scale measures the amount of transparency as perceived by the end user): *Low Control:* Users feel they can get stuck with this software, it can do unpredictable things, and they feel safer if they use only those parts of it they know will work. *High Control:* This software is easy to control and users can do exactly what they want with it. Users can get their work done with ease.

Learnability: This scale measures the ease with which a user can pick up how to use the software in the early learning stages, and the quality of the documentation, help files etc. *Low Learnability:* The software is initially very hard, and it's easy to forget how to do things. Users doubt they will ever master it. *High Learnability:* This

software is easy to get into, users could start working with it right away, and problems do not arise when a user needs to do something new.

Global: The global scale is not a subscale. While the subscales are based on 10 questionnaire items, the global scale is based on 25 items that pertain highly to the general usability dimension. These items are cross referenced in the subscales and contribute to their scores as well. According to Kirakowski "...when weighted according to the factor weights of the general usability scale and summed, these items make up the Global scale." (1998:37). The score on the *Global* scale is actually the weighted sum of the significant usability items in the SUMI scale. It is computed in such a way that, "...Global data is transformed into exactly the same scales as the five sub-scales, i.e. with a mean of 50 and a population standard deviation of 10." (Kirakowski 1998:37)

6.3.1.2 *Individual user scores*

Each user's views are documented in this output by SUMISCO. Any extreme view on a given scale is 'outlying' and lies outside the 'fences'. In this case the user is identified and marked A= Affect, C= Control etc. Table 2 in the SUMISCO report (section 6.4.2 and section 6.4.3) gives the individual scores. (Cf: SUMI reports in Appendix 6). These users can be later interviewed in order to probe for further information regarding their extreme opinions.

6.3.1.3 *Item Consensual Analysis (ICA)*

Kirakowski (1998:40) explains that "for each item in the inventory the program calculates the actual proportions of responses for the evaluation sample and the statistically expected proportions based on the standardization data for that item". In other words the collated and statistically prepared responses of the evaluators are matched to the standardized responses from the standardization data. This comparison helps the evaluator decide how much better or worse the evaluated software is from the generic standard.

6.3.1.3.1 *Statistical background*

The three answer columns in SUMI ('agree'/ 'undecided'/ 'disagree') are compared to the standardized tables in which the expected probabilities of responses in the standardization data are coded. These expected probabilities are multiplied by the number of users who are undertaking the evaluation in the sample being analysed to give the proportion of responses we would expect to find if the software was exactly like the standardisation data.

SUMISCO then computes the difference between the expected pattern of responses and what the respondents actually say. The statistic used is: $\chi^2 = \sum (f_o - f_e)^2 / f_e$. In this equation f_o is observed frequency of response and f_e is the expected frequency of response (Kirakowski 1998). Chi square distribution is used to ascertain whether the differences are due to chance variation or not. The statistic of SUMI scoring suggests that if there are lots of participating users more statistically different items will turn up, all other things being equal. In this case more stringent statistical criterion should be used. If there are fewer participants then fewer items will show up as significantly different and less stringent criterion can be used.

If the system or software is outstanding or poor then many more items will be marked as significant. For purposes of analysis and interpretation only those which are significantly different from the expected patterns of responses are selected, the ones which are not statistically different are not selected.

6.3.1.3.2 *Interpreting ICA*

After selecting the items for closer analysis they are separated from the rest of the file. Individual Chi square values are considered and the findings are related to what is known about the context of use and context of test. Correlations are drawn and phenomenon explained in the light of these. Close attention needs to be paid to items where the 'undecideds' predominate as this could be indicative of three things: 1) the item is not applicable to the program; 2) the user desires to express a milder verdict; and 3) the user is really not decided in their mind. Equipped with the knowledge of context of use decisions can be made about whether the users really want to give a milder verdict, and, if this is the case, which direction the verdict leans towards. This can be determined by looking at the rest of the items that this item clusters with, and by determining whether there are higher chi square weights for 'agrees' or 'disagrees' after eliminating the 'undecideds'.

6.3.2 *SUMI results for Listening to lectures*

In this section the scoring report of EASE Vol 1(Day2) *Listening to Lectures* is presented with my comments appearing in italics. The whole report can be seen in Appendix of Chapter 6. Graphs are presented of selected tables and significant items consensual analysis. Since SUMI was administered twice DAY 1's results are not given here in detail only the *Profile Analysis* is given at the end of the section. Kirakowski (1998) recommends that the results from SUMI be presented in the way they are presented below (for the first EASE Volume *Listening to Lectures*). He also recommends the use of italics for the author's (mine) commentary as apposed to the

normal font of the report's content and notes to distinguish between the two. However his recommended style is not followed entirely after the first instance (SUMI Scoring Report SUMISCO 7.38 for *EASE Listening to Lectures DAY2*). The use of italics continues for my commentary but for explanation of how to read the *Profile Analysis* and *Item Consensual Analysis* for subsequent reporting of results the reader is referred back to read the explanation given in the immediately following section.

SUMI Scoring Report from SUMISCO 7.38

Time and date of analysis: 11:34:03 on 04-04-2006

Files used in this analysis:

SUMI English (UK) Language Items

SUMI Version 2.1 Scoring Keys

distributions: set 01A

weights: set 01A

population parameters: set 01A

Data file analysed: ease1d2.ASC: day two EASE Vol1: Listening to lectures

Number of users analysed: 41

Profile Analysis

Scale	UF	Ucl	Medn	Lcl	LF
Global	75	61	58	55	42
Efficiency	77	63	59	55	41
Affect	73	56	53	50	28
Helpfulness	73	65	62	59	49
Control	65	57	54	51	41
Learnability	73	65	61	57	49

Note:

The Median is the middle score when the scores are arranged in numerical order. It is the indicative sample statistic for each usability scale.

The Ucl and Lcl are the Upper and Lower Confidence Limits. They represent the limits within which the theoretical true score lies 95% of the time for this sample of users.

The UF and LF are the Upper and Lower Fences. They represent values beyond which it may be plausibly suspected that a user is not responding with the rest of the group: the user may be responding with an outlier.

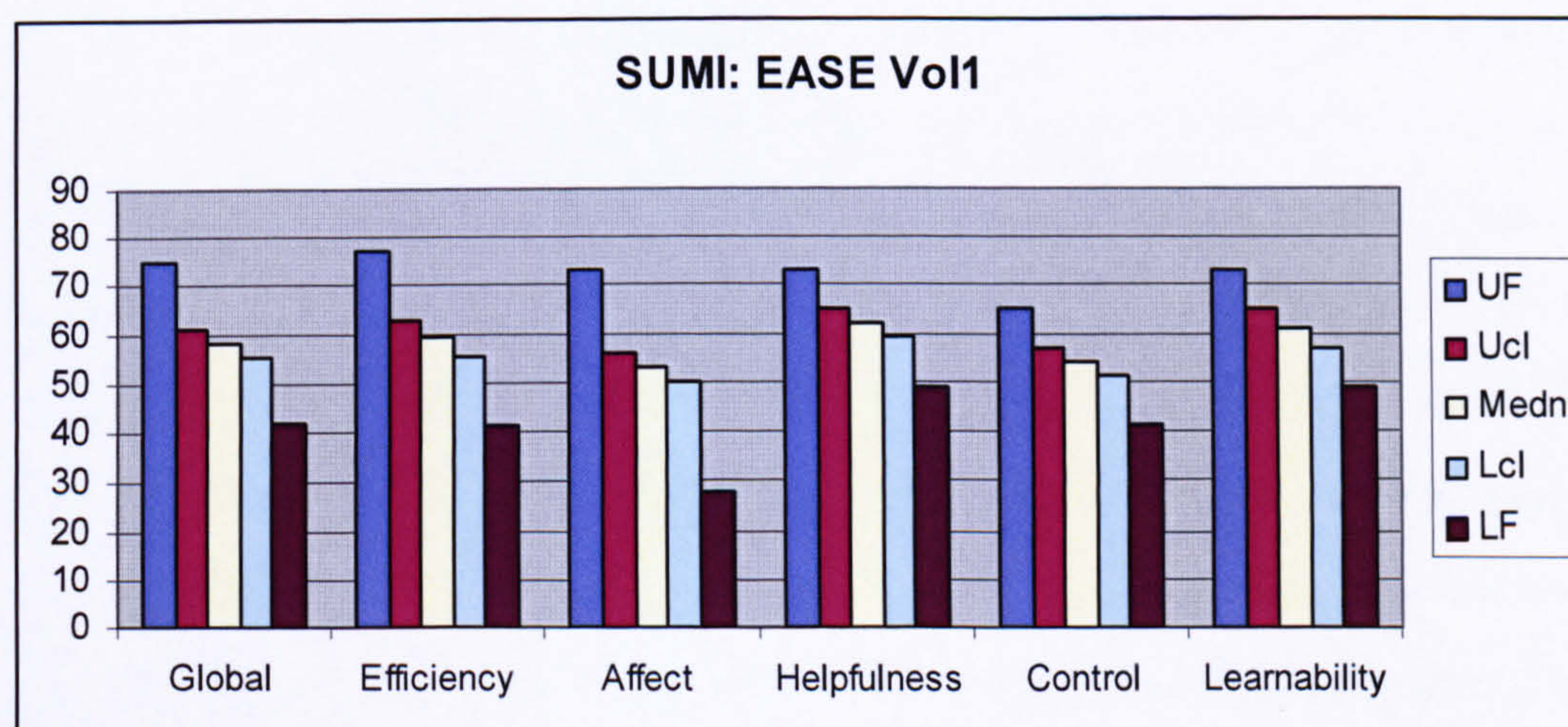


Figure 6.3 SUMI Results of *Listening to Lectures*

A score of above 50 is deemed acceptable quality of usability and a score of 60 is considered good (Kirakowski1998). The scale of Helpfulness (median score 62) and Learnability (median score 61) are at the higher end of the rating. The lowest ranking is Affect (median score 53) or likeability subscale. Global Usability is at 58 which is a good result for the software, although the authors of SUMI suggest that further study of Individual user scores and Item Consensual Analysis can always shed further light on areas for Software improvement.

Individual User Scores

User	Globa	Effic	Affec	Helpf	Contr	Learn	Identifier
1	61	60	54	64	60	60	001
2	61	71	47	65	52	64	002
3	64	63	66	67	50	61	003
4	71	71	65	67	67	64	004 (C)
5	55	65	54	57	53	63	005
6	57	62	43	56	53	61	006
7	66	63	68	64	57	62	007
8	48	57	22	55	56	68	008 (A)
9	23	13	41	22	30	23	009 (GEHCL)
10	64	64	66	62	68	63	010(C)
11	56	65	41	60	58	69	011
12	30	28	34	31	32	13	012 (GEHCL)

If we look at the first 12(out of 41) users individual scores we see that user 004 is an outlier in the subscale of Control and user 008 on the subscale of Affect. These are individual low ratings; what is significant is a low rating throughout the subscales rather than in just one aspect of the program. Users Q-009 and Q- 012 are outliers on five out of six (GEHCL) scales. The implications of this

are that these two students gave low rating to all aspects of the program except Affect. This is a surprising finding because if the program scored poorly on all other aspects it is difficult to see how it could be likeable. To rule out any possibility of error in ticking the right/wrong box because the questionnaire format that I used did not first of all align the options of 'agree', 'undecided' and 'disagree' close to the boxes to be ticked. Secondly the format of the questionnaire did not repeat the options of 'agree', 'undecided' and 'disagree' after every item. Moreover, it was a multi page questionnaire, (there was a chance that the respondent had forgotten or misunderstood which box represented what,) therefore, I went back to check the original actual complete questionnaires filled out by these students. What I found was that the middle box of 'undecided' was ticked quite volitionally for items that accounted for positive elements of the software, thus establishing that these respondents actually were unsure about the software. User Q-009 could have ticked the wrong boxes deliberately or could have made an inadvertent error; this outlier view is not supported by the other questionnaire data that I have from him. The possibility of such an error occurring cannot be ruled out entirely but since I have 41 respondents, outliers can be well tolerated by the analysis. Another option available was to look at the PLUM questionnaire results for these two students to establish the authenticity of this data. User Q-009 had not participated in the PLUM questionnaire evaluation but User Q-012's negative evaluation is corroborated by this respondent's responses to similar items in the PLUM questionnaire. However, since PLUM is a differently designed questionnaire which does not have five subscales and a Global usability scale, any conclusion from this comparison must be tentative.

Item Consensual Analysis

In the following table, the numbers in the row labelled 'Profile' are the observed responses of the actual users to each item.

The numbers in the row labelled 'Expected' are the number of responses expected on the basis of the standardisation database.

The Goodness of Fit between the observed and expected values is summarised using Chi Square, and these statistics are presented on the line below the expected values.

The number at the end of the Goodness of Fit line is the total Chi Square which applies to that item. The greater the value of the total Chi Square, the more likely it is that the obtained values differ from what is expected from the standardisation database.

Each total Chi Square marked with

- *** is at least 99.99% certain to be different
- ** is at least 99% certain to be different
- * is at least 95% certain to be different

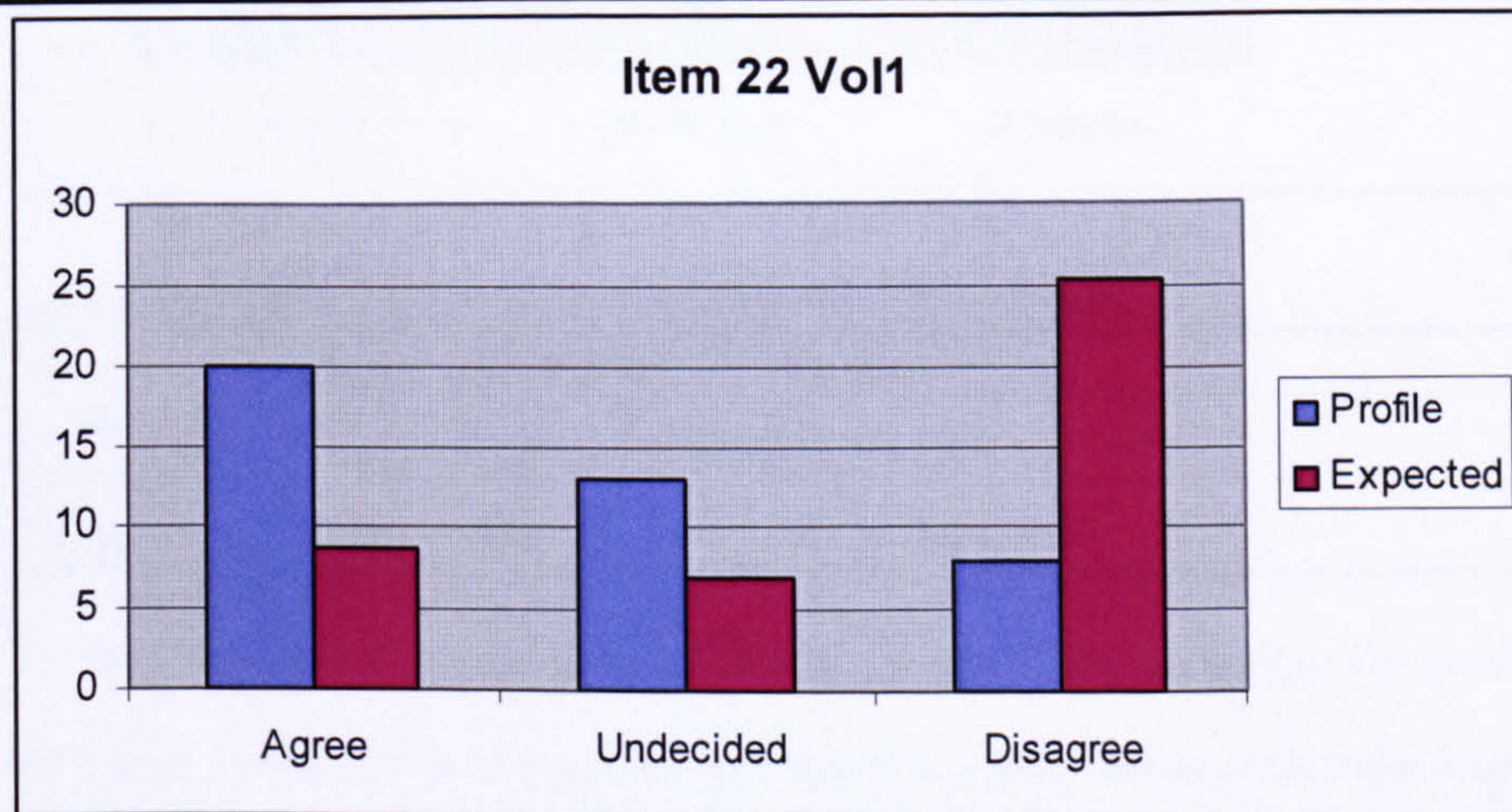
Total Chi Square values without asterisks are not likely to differ much from the standardisation database.

In this output, the SUMI items which differ most from the standardisation are presented first.

Only a subset of the ICA items are shown here: graphs are drawn for commented items.

I would not like to use this software every day.

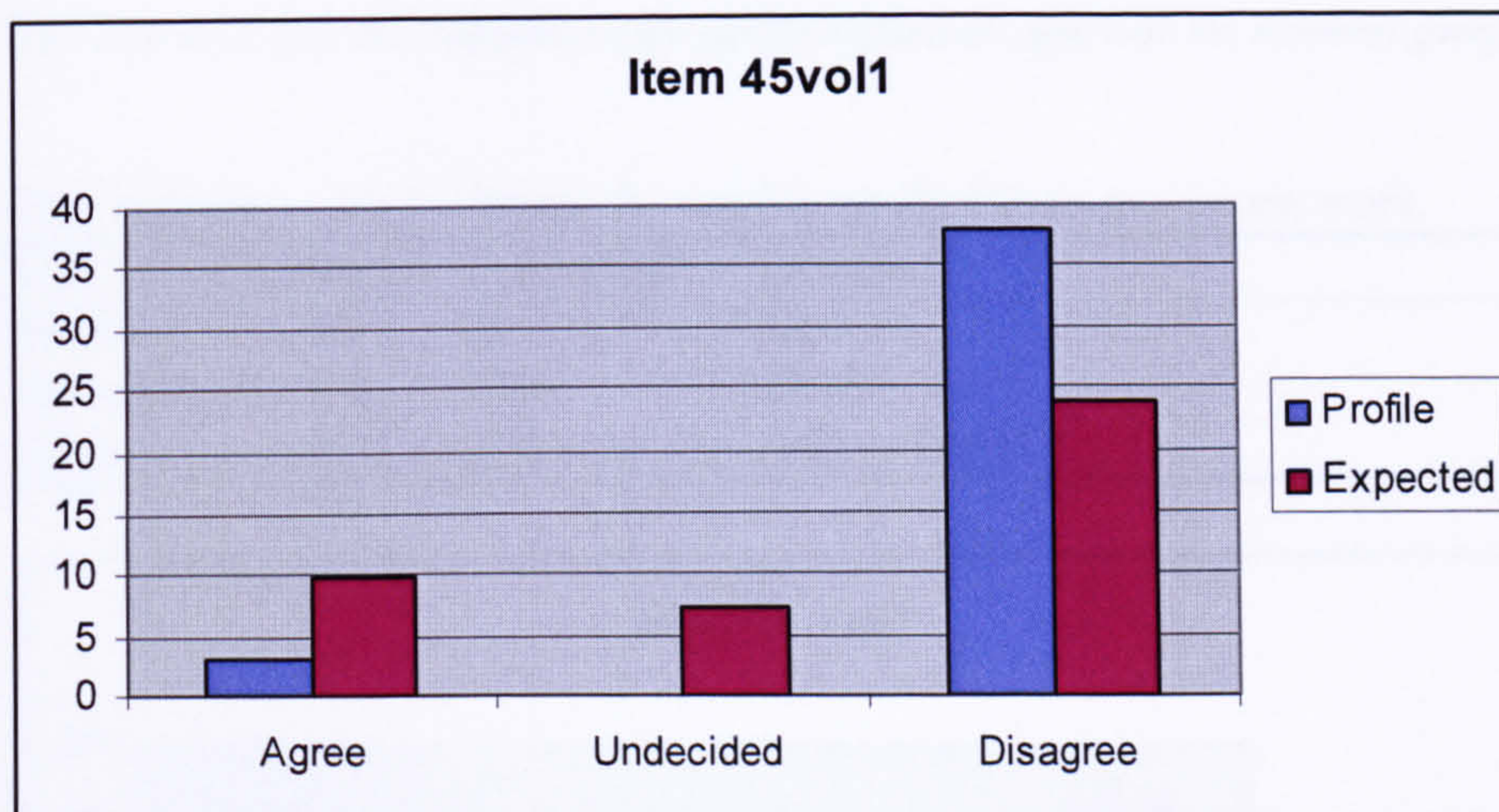
Item 22	Agree	Undecided	Disagree
Profile	20	13	8
Expected	8.8	6.91	25.29
Chi Sq	14.24	5.38	11.82
			31.44***



The results of item 22 show that the Chi square difference is highest for the 'agree' answer. Half the respondents indicate that they would not like to use this software every day. This difference from expected scores was expected because of the nature of the program being evaluated.

It is easy to forget how to do things with this software.

Item 45	Agree	Undecided	Disagree
Profile	3	0	38
Expected	9.89	7.28	23.83
Chi Sq	4.8	7.28	8.42
			20.49***



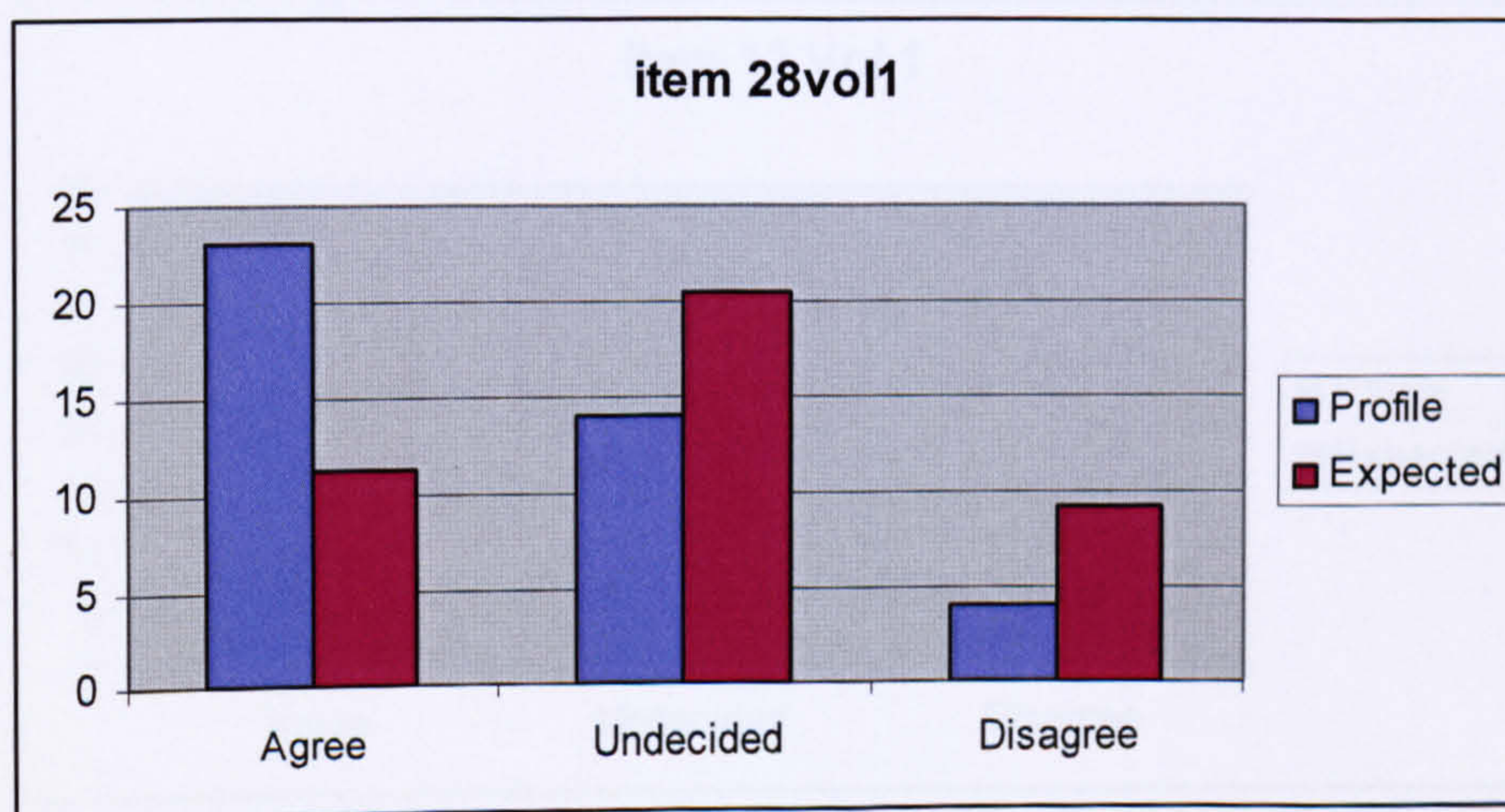
I sometimes don't know what to do next with this software.

Item 6	Agree	Undecided	Disagree
Profile	2	5	34
Expected	13.14	7.08	20.79
Chi Sq	9.44	0.61	8.4
			18.45***

The statement for item 45 is a negative one so more people disagree then agree. The biggest Chi Square difference is contributed by the 'undecideds'. This significance is more positive then negative as a greater number in the profile disagrees with the statement then the expected. The same is the case with 6.

The software has helped me overcome any problems I have had in using it.

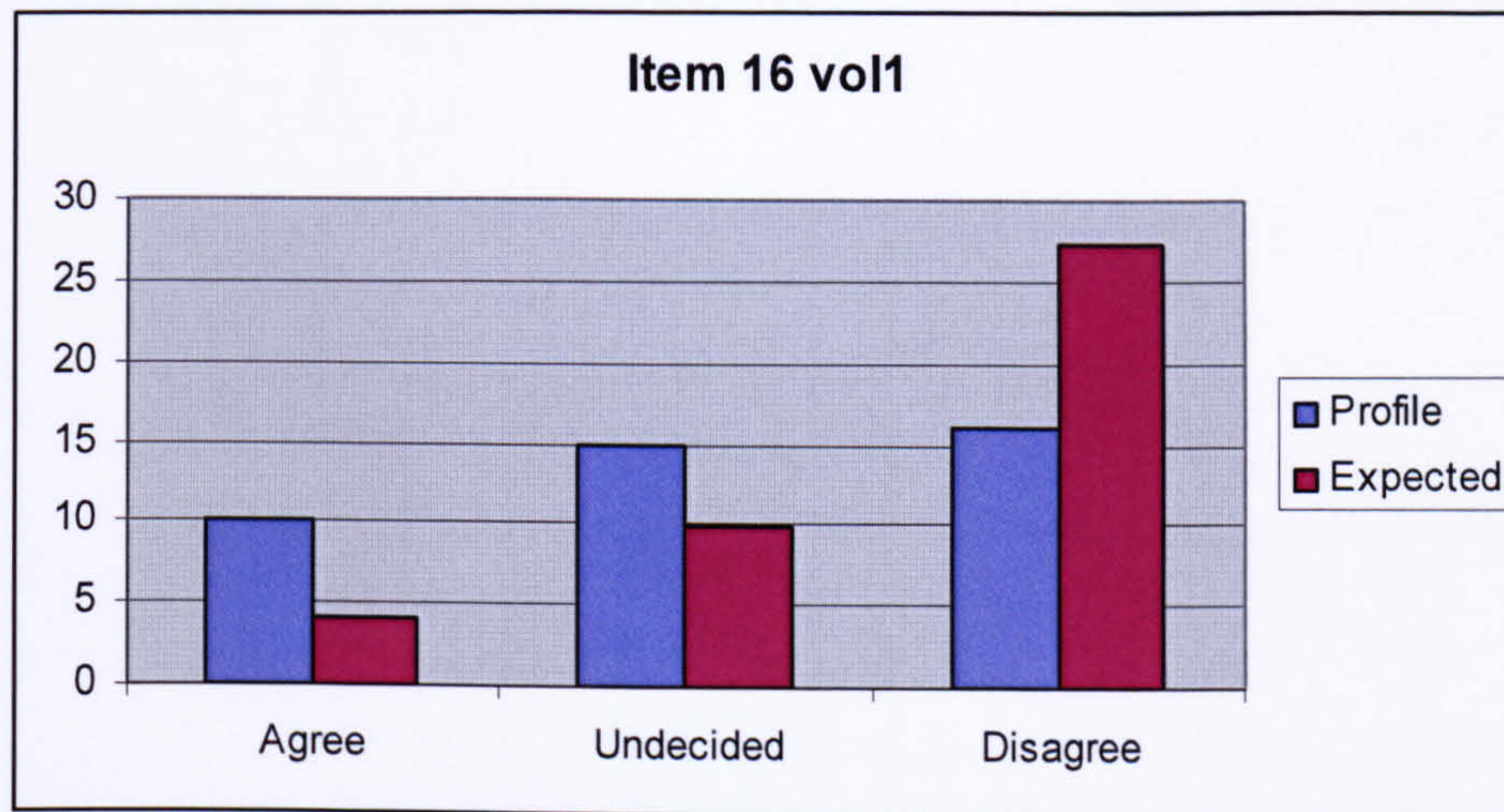
Item 28	Agree	Undecided	Disagree
Profile	23	14	4
Expected	11.31	20.48	9.21
Chi Sq	12.09	2.05	2.95
			17.09***



This item deals with the Helpfulness scale and the respondents agree with the statement giving it a positive rating.

This software seems to disrupt the way I normally like to arrange my work.

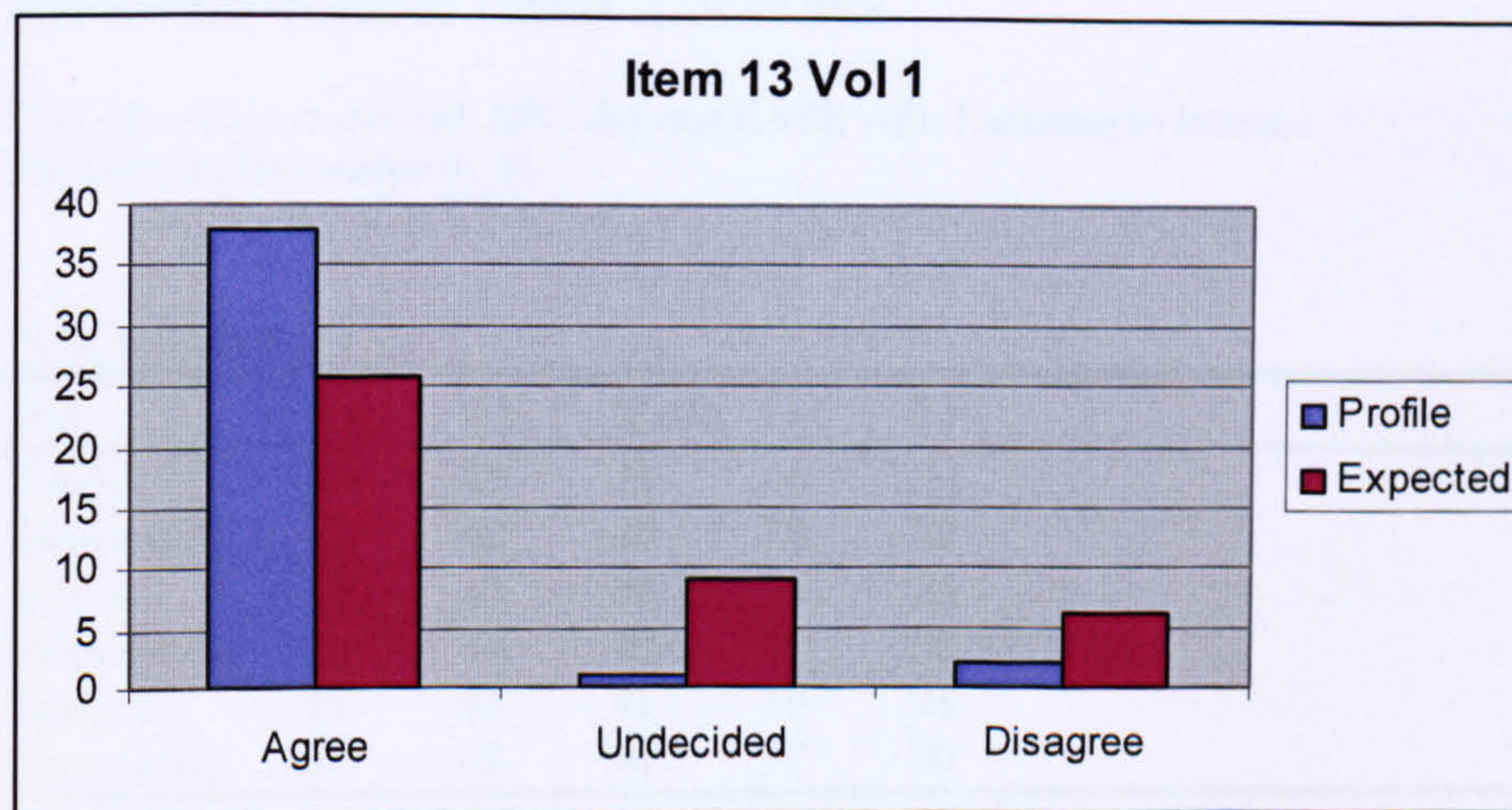
Item 16	Agree	Undecided	Disagree
Profile	10	15	16
Expected	4.0	9.78	27.22
Chi Sq	9.03	2.78	4.63
			16.43***



This answer is quite close to the expected the highest chi square coming in responses agreeing with the statement.

The way that system information is presented is clear and understandable.

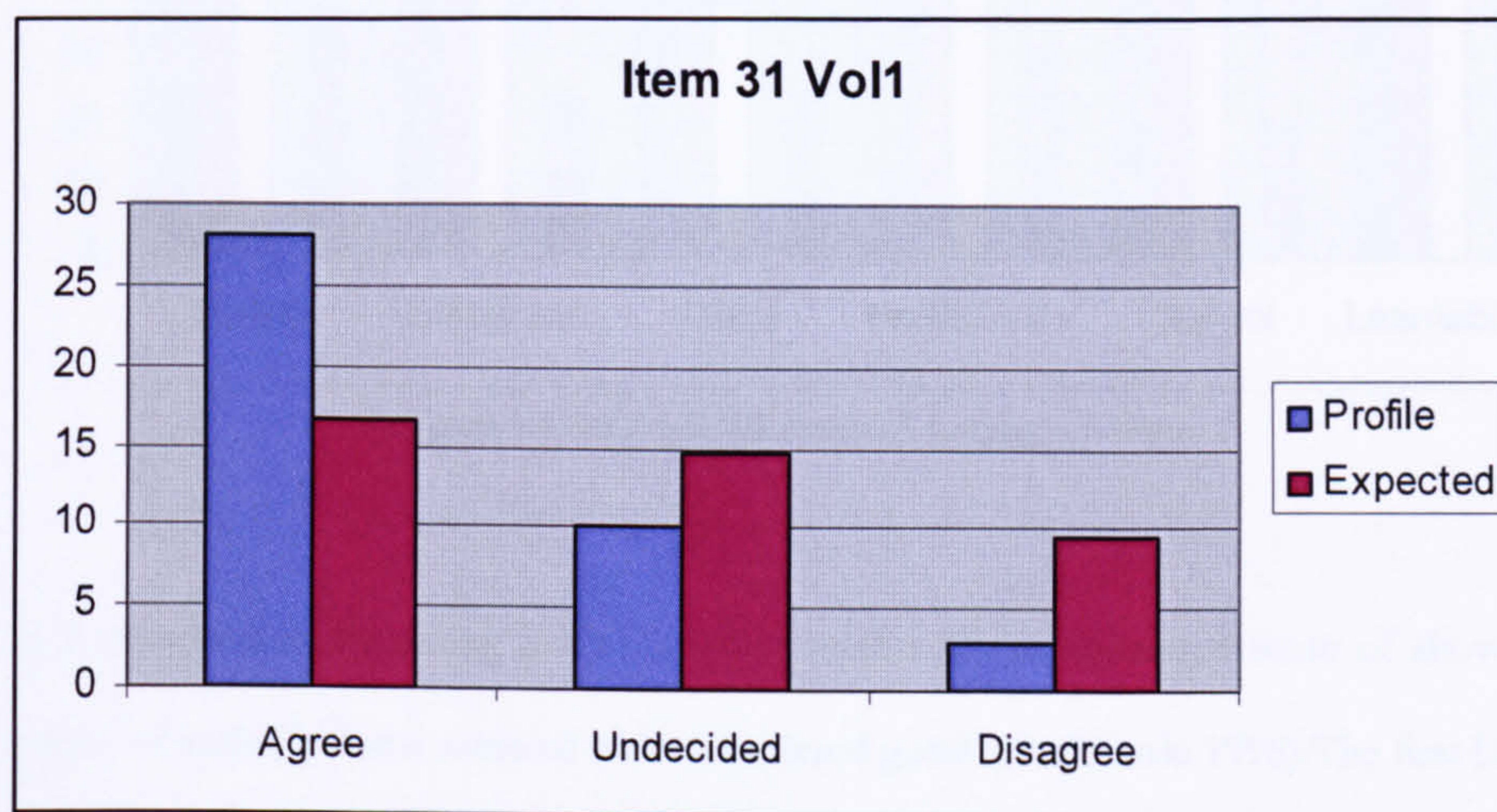
Item 13	Agree	Undecided	Disagree
Profile	38	1	2
Expected	25.83	9.01	6.16
Chi Sq	5.73	7.12	2.81
			15.66***



This item is close to the expected results with the greatest difference in chi square presented by the 'undecideds'. The profile results on 'agree' are higher than the expected so the significance is positive.

It is obvious that user needs have been fully taken into consideration.

Item 31	Agree	Undecided	Disagree	
Profile	28	10	3	
Expected	16.62	14.8	9.58	
Chi Sq	7.79	1.55	4.52	13.86***



This item also has a higher profile score for 'agree' than expected scores, indicating that the significance is positive.

6.3.2.1 DAY One SUMI results for EASE Volume One Listening to Lectures

The Results from DAY 1 of SUMI evaluation of Listening to Lectures are given in brief to show the discrepancy in results.

SUMI Scoring Report from SUMISCO 7.38

Time and date of analysis: 11:33:55 on 04-04-2006

...

Data file analysed: ease1d1.ASC: day one EASE vol1: Listening to lectures

Number of users analysed: 36

Profile Analysis

Scale	UF	Ucl	Medn	Lcl	LF
Global	48	42	41	40	34
Efficiency	53	42	40	38	24
Affect	49	40	38	36	25
Helpfulness	53	44	42	40	32
Control	61	43	41	39	21
Learnability	57	42	40	37	20

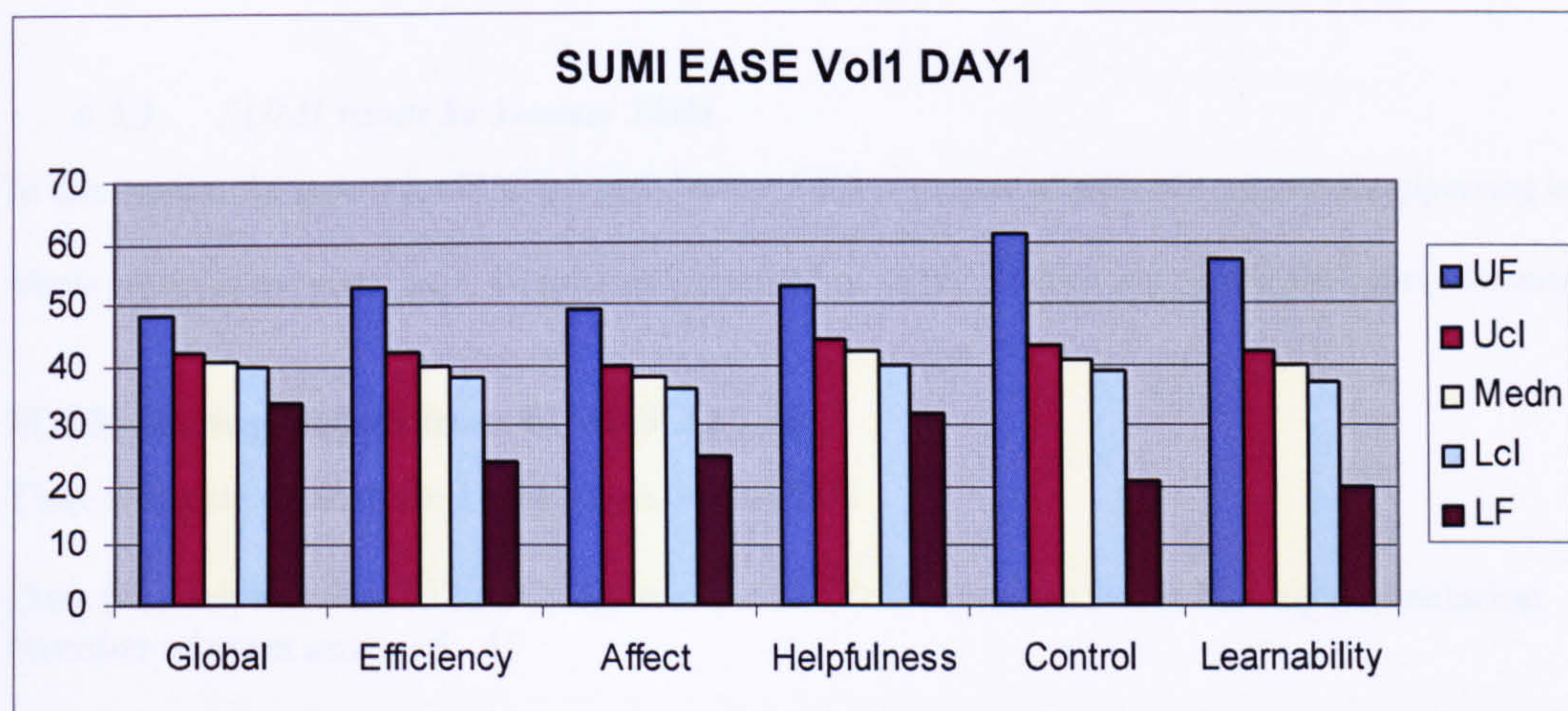


Figure 6.4 DAY-1 SUMI results of *Listening to Lectures*

As I stated when discussing DAY 2's results earlier in this section: a score of above 50 is deemed acceptable quality of usability and a score of 60 is considered good (Kirakowski 1998). The first DAY 1 evaluation of EASE materials yielded considerably lower scores. The scale of Helpfulness (median score 42) and Control and Global scales (median scores 40) are at the higher end of the rating. The lowest ranking is Affect (median score 38) or likeability subscale.

Table 6.3 Comparison of Day 1 and Day 2 Results for *Listening to Lectures*

	day1	day2
Scale	Medn	Medn
Global	41	58
Efficiency	40	59
Affect	38	53
Helpfulness	42	62
Control	41	54
Learnability	40	61

The two results vary by a difference of almost 20 points but the highest rated scale is Helpfulness in both days results (day 1 median score: 42, day 2 median score: 62). The lowest rated is Affect on both days (Day1= 38, Day 2=53)

6.3.3 SUMI results for Seminar Skills

In this section the report for EASE Vol 2 *Seminar Skills* is presented with my comments appearing in italics. The whole report is not given here. Graphs are presented of selected tables and significant items consensual analysis.

SUMI Scoring Report from SUMISCO 7.38

Time and date of analysis: 11:34:15 on 04-04-2006

...

Data file analysed: ease2d2.ASC: day two EASE Vol2: Seminar Skills: Making presentation

Number of users analysed: 41

Profile Analysis

Scale	UF	Ucl	Medn	Lcl	LF
Global	83	67	63	59	35
Efficiency	84	67	63	59	36
Affect	79	63	60	57	34
Helpfulness	76	68	64	60	46
Control	73	60	57	54	37
Learnability	75	66	62	58	48

Refer to section 6.4.2 Profile Analysis guidance note for explanation of the above table.

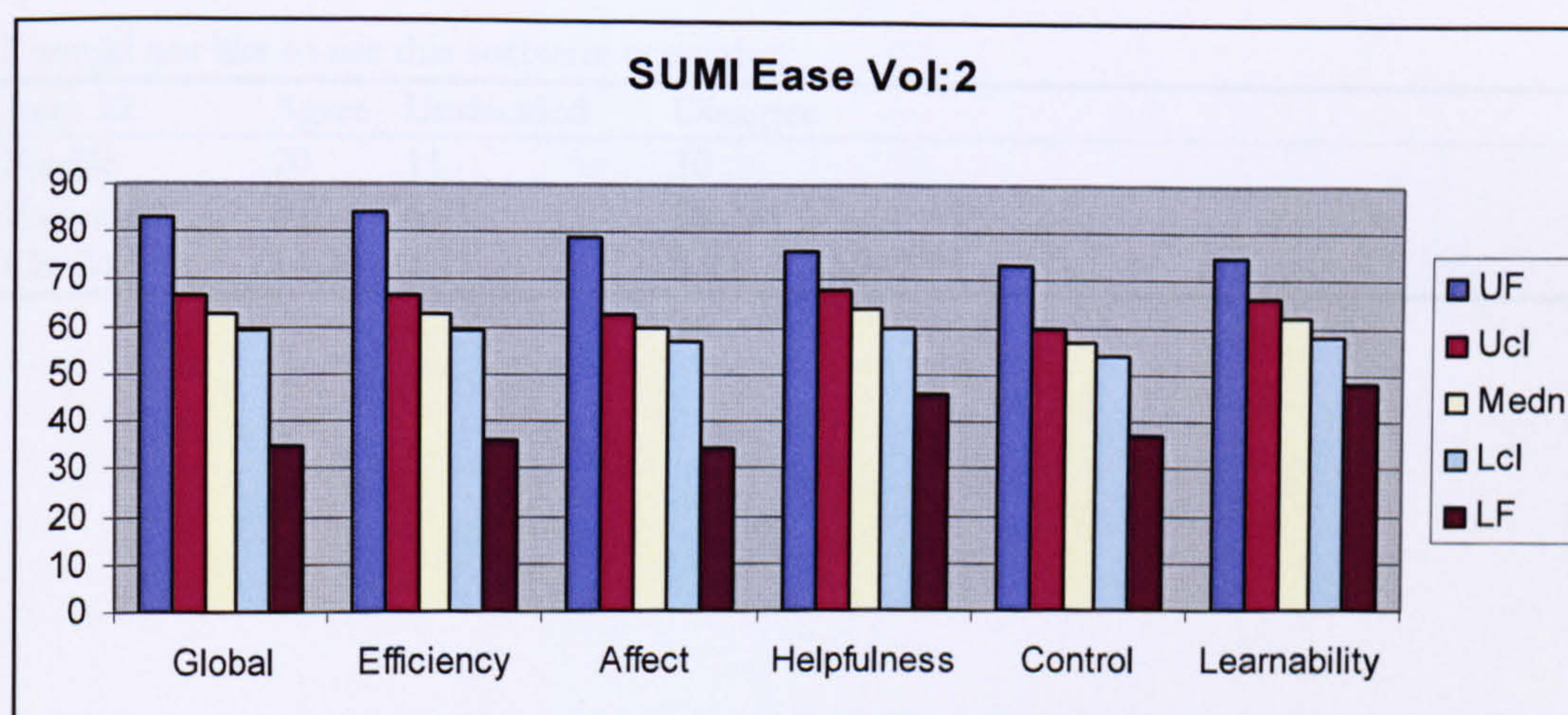


Figure 6.5 SUMI results of *Seminar Skills*

Individual User Scores

User	Globa	Effic	Affec	Helpf	Contr	Learn	Identifier
1	48	66	40	52	42	57	Q-001
2	65	71	60	65	59	64	Q-002
3	66	65	66	69	56	59	Q-003
4	71	71	65	65	67	62	Q-004
5	53	56	54	57	53	63	Q-005
6	63	71	60	65	59	60	Q-006
7	66	63	68	64	56	66	Q-007
8	41	50	19	47	57	60	Q-008 (A)
9	23	13	41	22	30	23	Q-009(GEHCL)
10	69	64	65	69	68	66	Q-010
11	67	71	63	66	64	71	Q-011
12	33	38	40	25	31	19	Q-012 (GHCL)

It is not surprising to note that it is the same respondents who are the outliers again but to a lesser degree than for volume one. This is supported by the general mood of the evaluation; students liked volume two better. However, this does put to rest my earlier concerns that there could be other unaccounted reasons for outlying scores. It appears that these respondents are consciously and volitionally outliers.

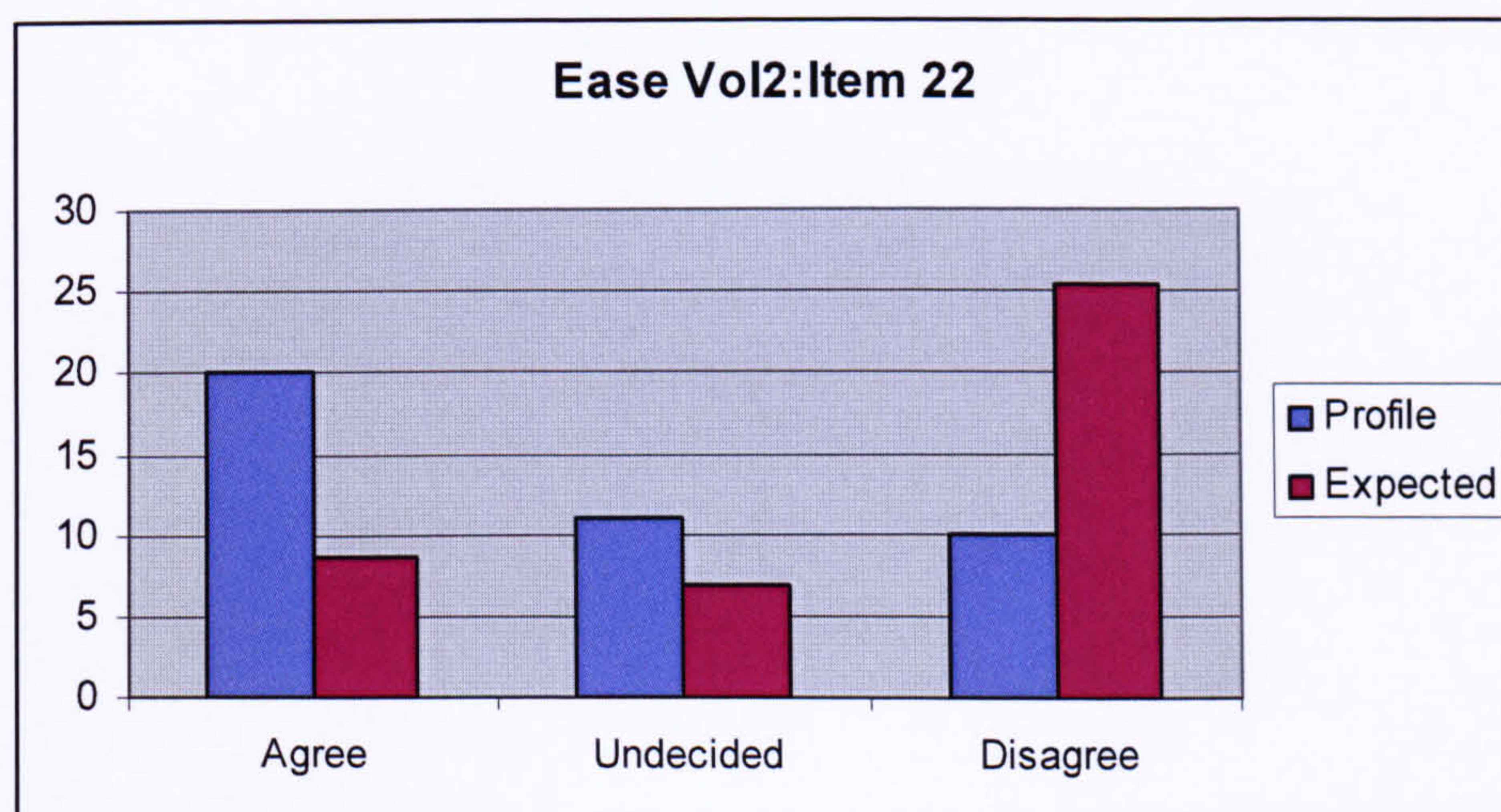
Item Consensual Analysis

Refer to Section 6.4.2 for guidance notes on how to read the results given below or see further Appendix 6 SUMISCO reports on EASE Vol:1 and Vol:2

Only a subset of the ICA items are shown here: graphs are drawn for the items singled out to be shown here from the whole 50 items in the SUMISCO report(Cf: Appendix 6 SUMISCO reports.

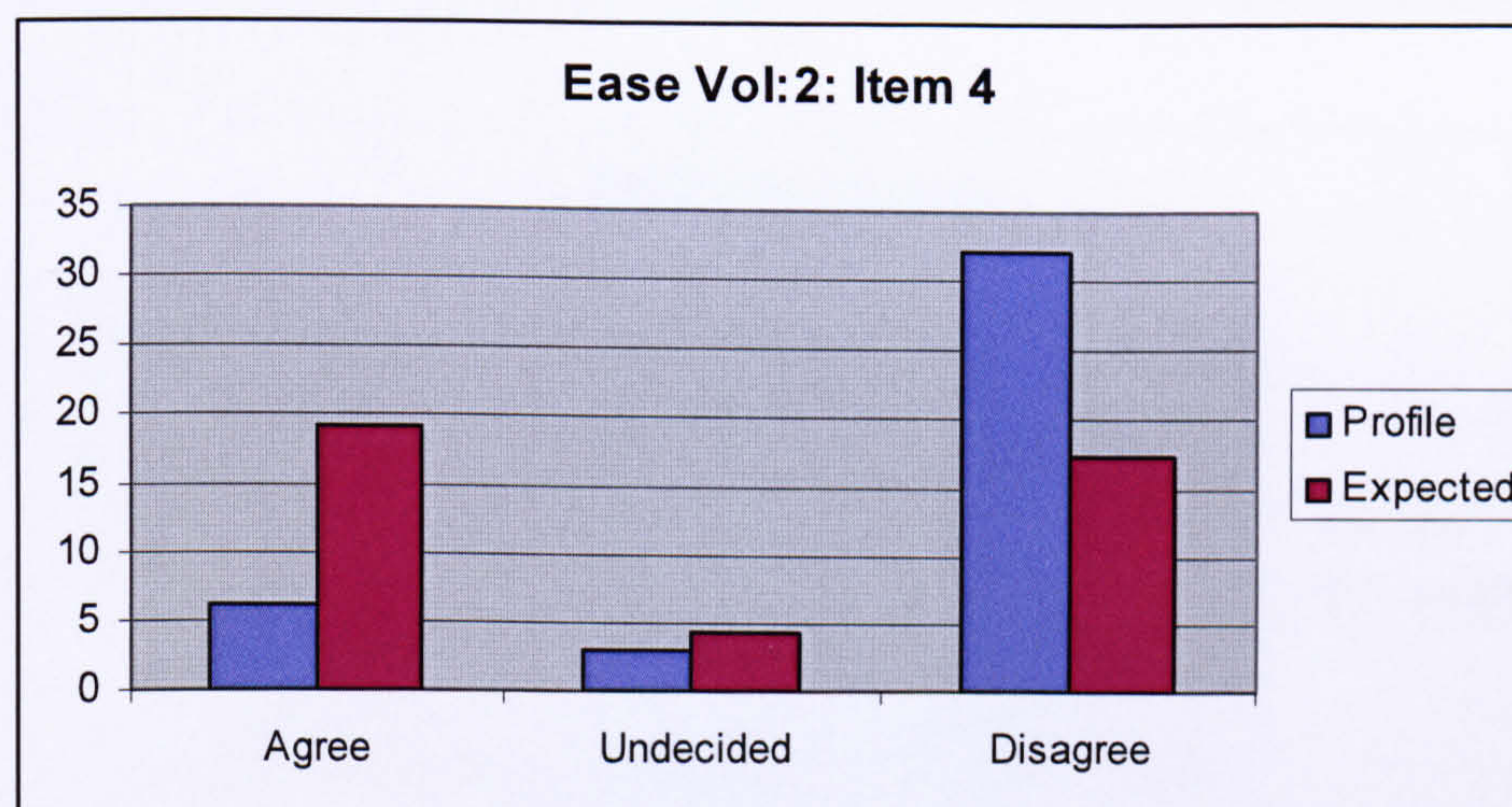
I would not like to use this software every day.

Item 22	Agree	Undecided	Disagree
Profile	20	11	10
Expected	8.8	6.91	25.29
Chi Sq	14.24	2.43	9.24 25.91***



The software has at some time stopped unexpectedly.

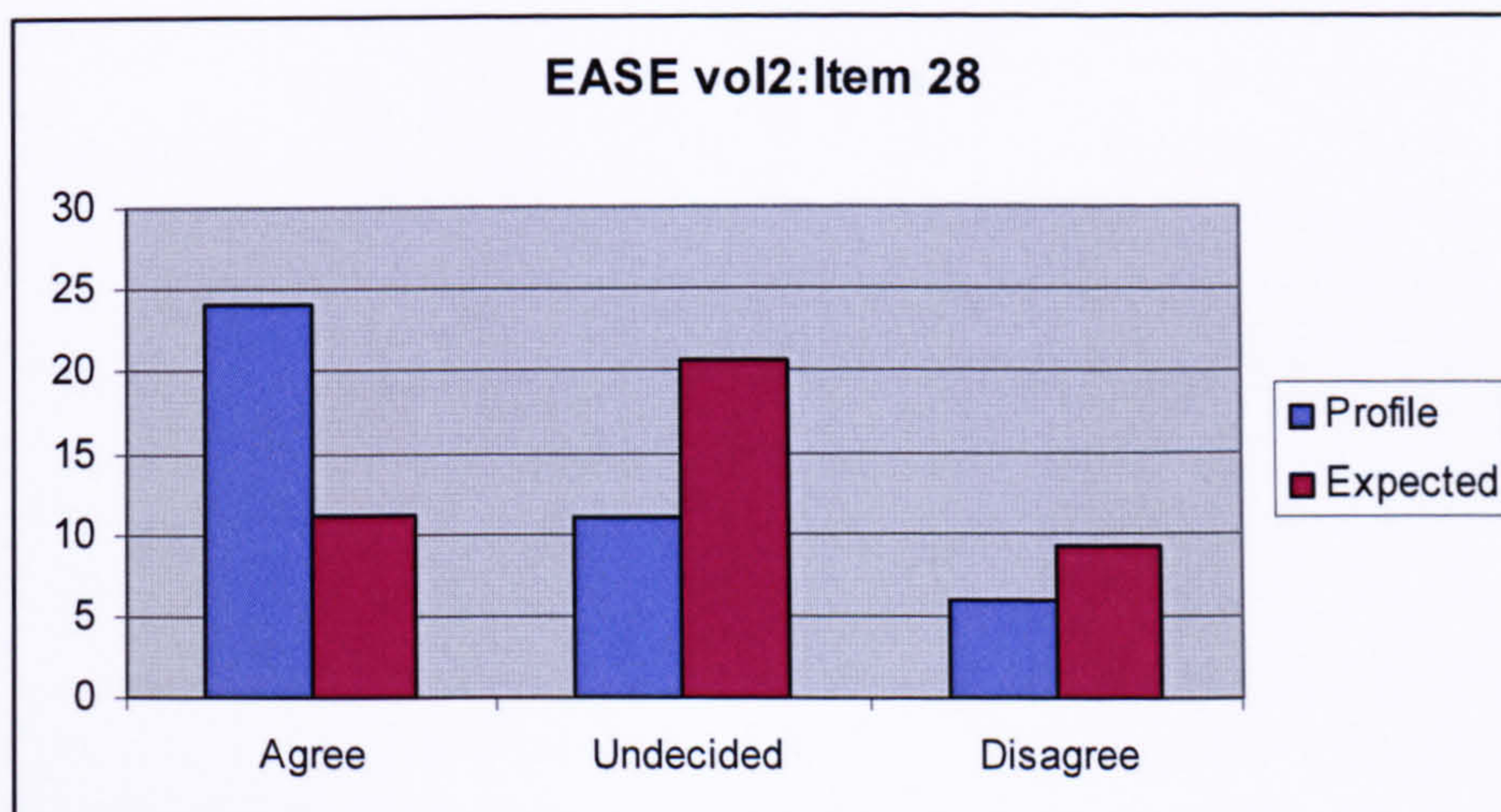
Item 4	Agree	Undecided	Disagree
Profile	6	3	32
Expected	19.23	4.33	17.44
Chi Sq	9.1	0.41	12.17
			21.68***



This difference is a positive rating to the software. More respondents in my study disagree with the statement than the standardized data and fewer agree than the expected.

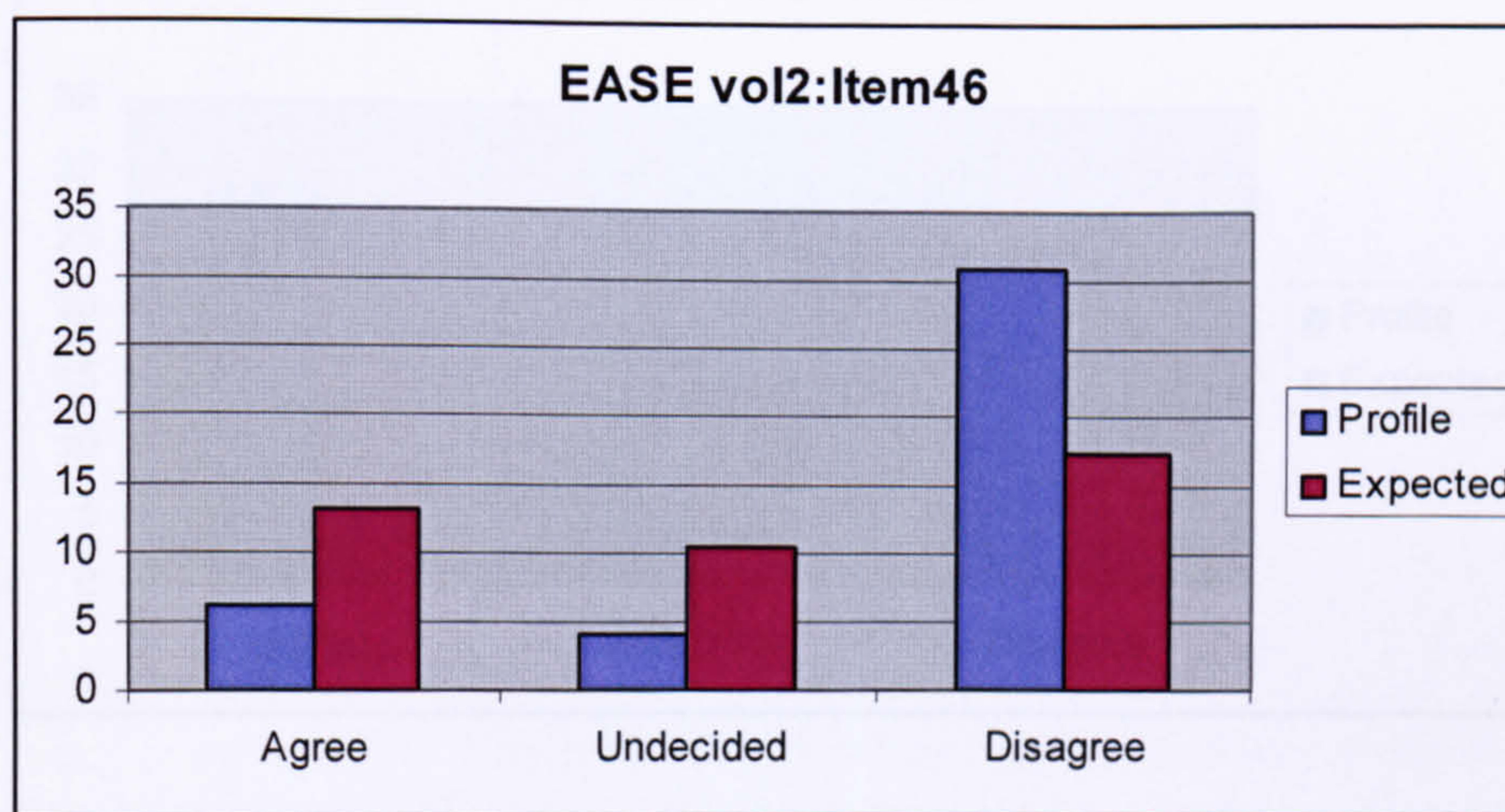
The software has helped me overcome any problems I have had in using it.

Item 28	Agree	Undecided	Disagree
Profile	24	11	6
Expected	11.31	20.48	9.21
Chi Sq	14.25	4.39	1.12
			19.75***



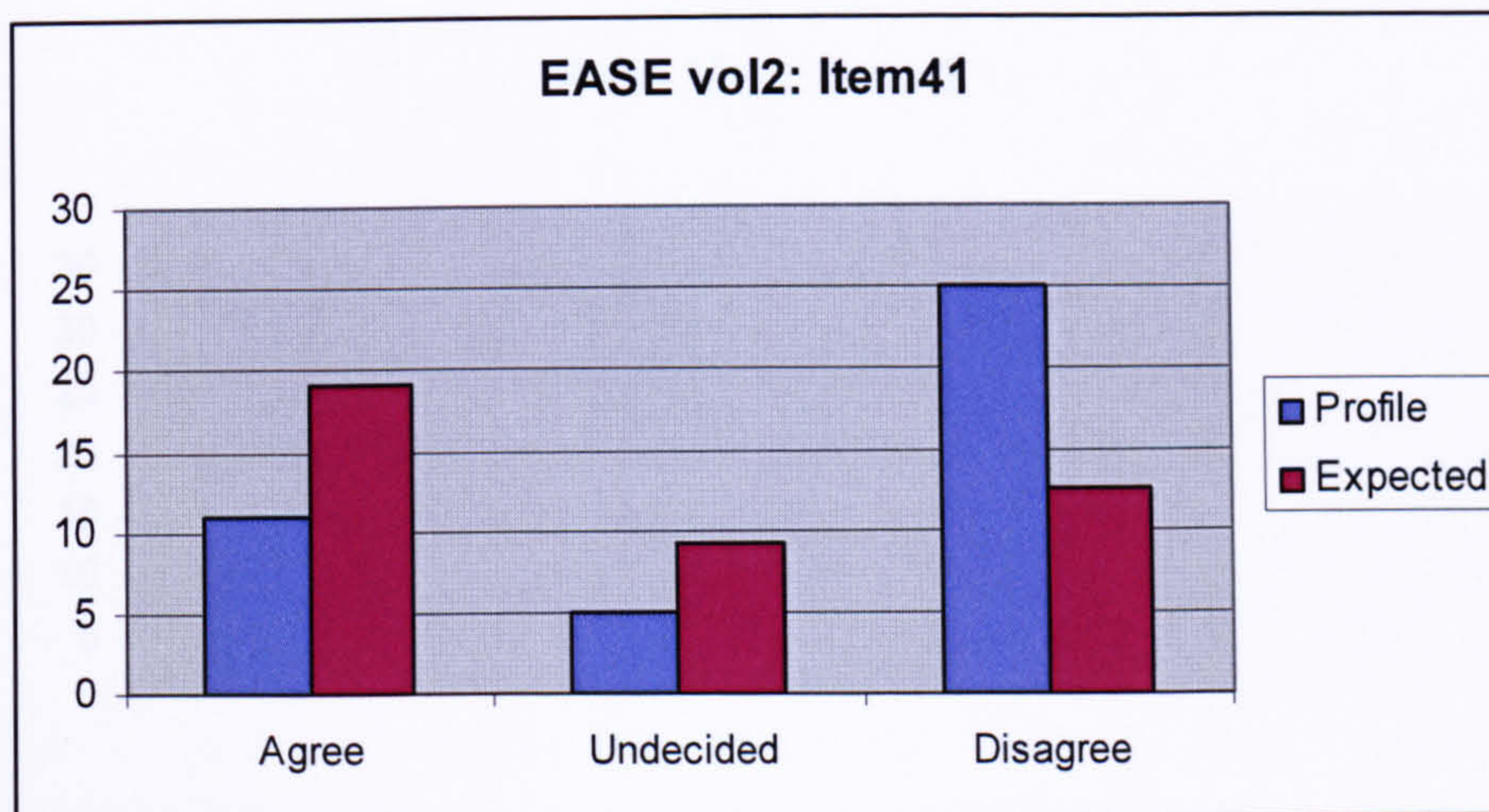
This software occasionally behaves in a way which can't be understood.

Item 46	Agree	Undecided	Disagree	
Profile	6	4	31	
Expected	13.2	10.33	17.47	
Chi Sq	3.93	3.88	10.48	18.28***



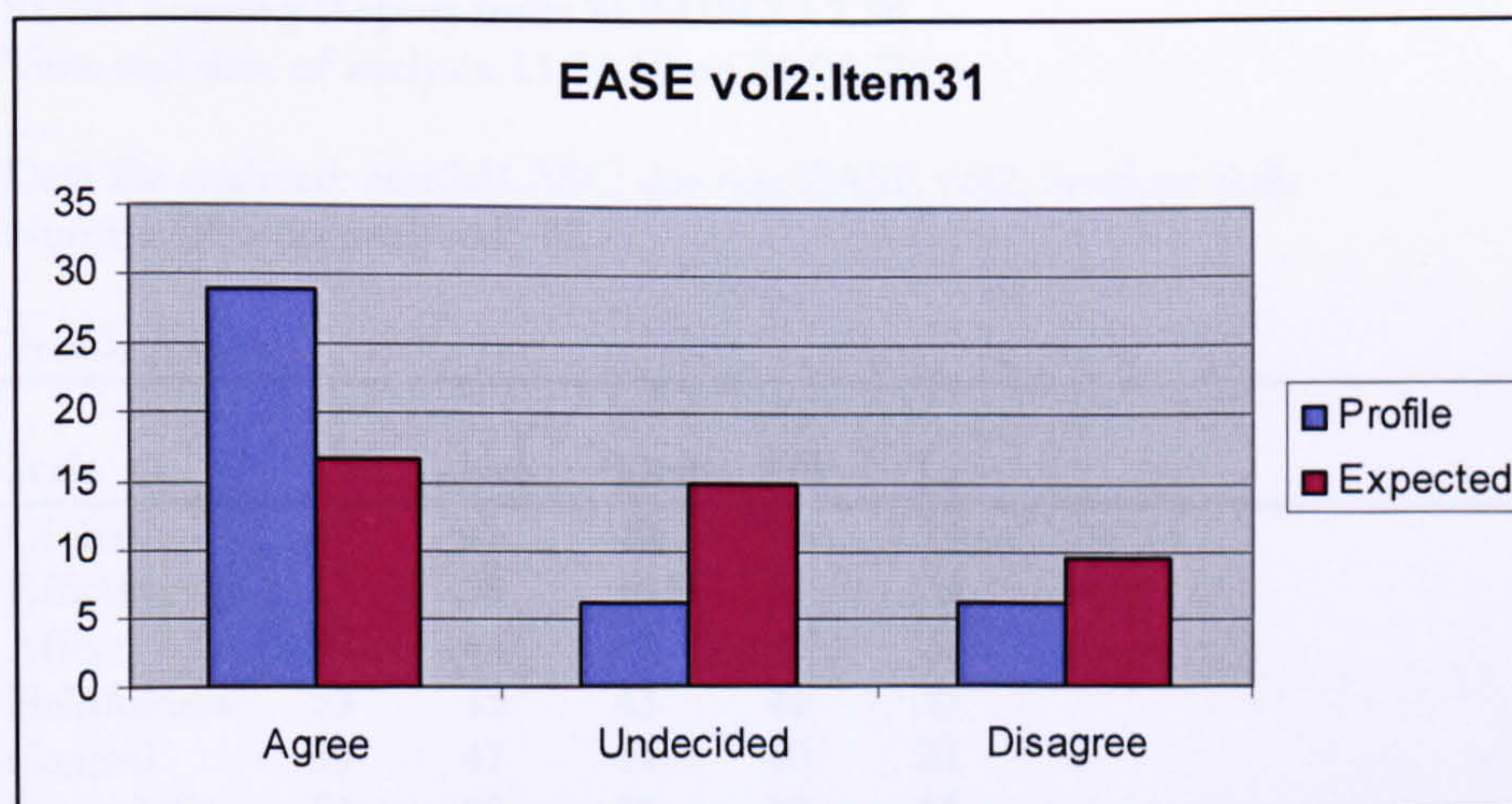
The software hasn't always done what I was expecting.

Item 41	Agree	Undecided	Disagree	
Profile	11	5	25	
Expected	19.06	9.28	12.66	
Chi Sq	3.41	1.97	12.02	17.4***



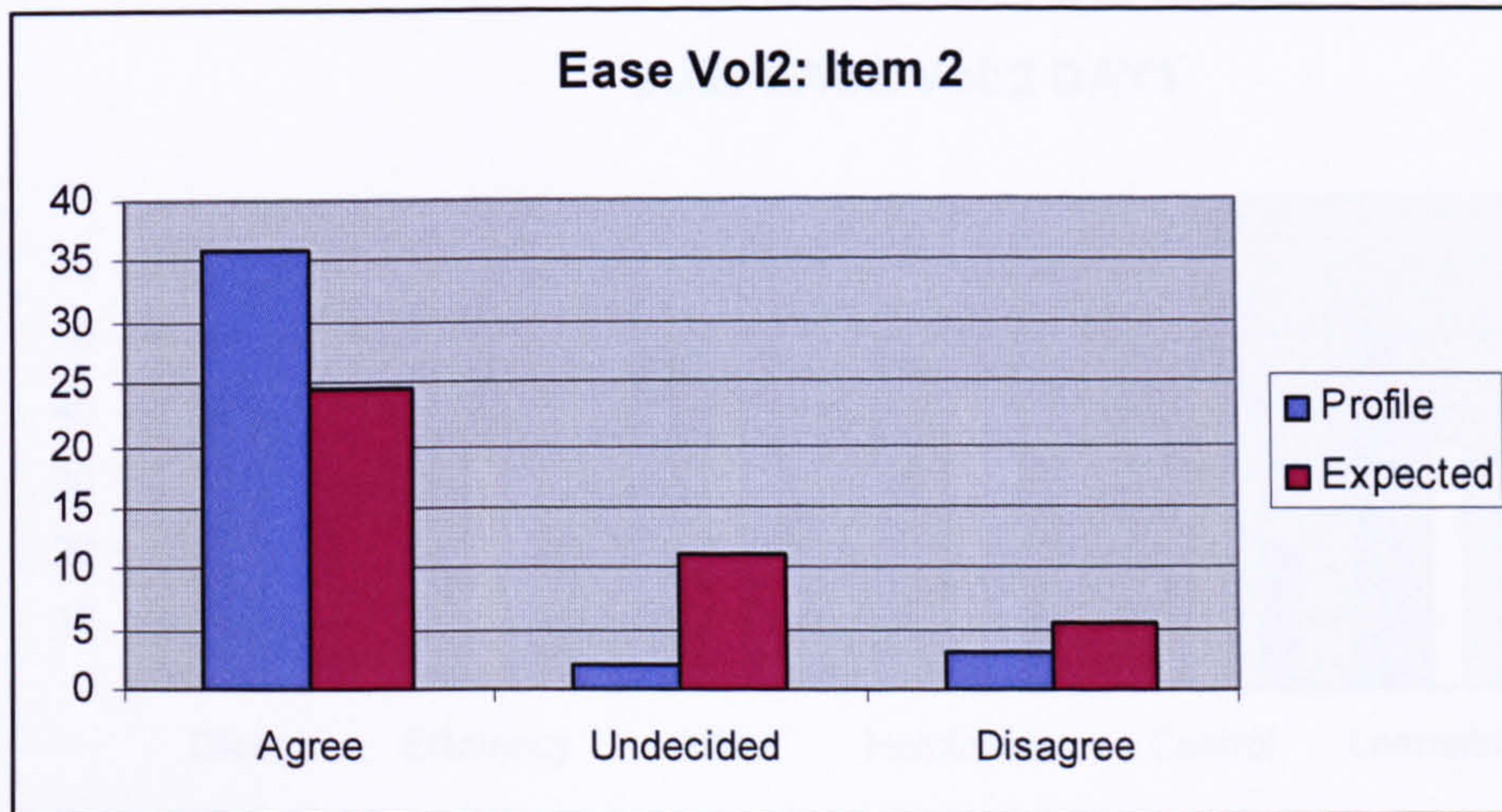
It is obvious that user needs have been fully taken into consideration.

Item 31	Agree	Undecided	Disagree	
Profile	29	6	6	
Expected	16.62	14.8	9.58	
Chi Sq	9.21	5.23	1.34	15.78***



I would recommend this software to my colleagues.

Item 2	Agree	Undecided	Disagree	
Profile	36	2	3	
Expected	24.51	10.97	5.52	
Chi Sq	5.38	7.33	1.15	13.87***



6.3.3.1 DAY One SUMI results for EASE Volume Two: Seminar Skills

SUMI results of EASE Vol 2 *Seminar Skills* on day one were lower than on Day 2 which were given above. Only the profile analysis is given here to give an idea of the scale of difference between these two evaluations.

SUMI Scoring Report from SUMISCO 7.38

Time and date of analysis: 11:34:10 on 04-04-2006

...

Data file analysed: ease2d1.ASC: day one EASE vol2: Seminar skills

Number of users analysed: 32

Profile Analysis

Scale	UF	Ucl	Medn	Lcl	LF
Global	47	41	40	39	33
Efficiency	47	38	36	34	23
Affect	57	44	42	40	27
Helpfulness	53	45	43	41	33
Control	59	47	44	40	21
Learnability	52	38	35	32	15

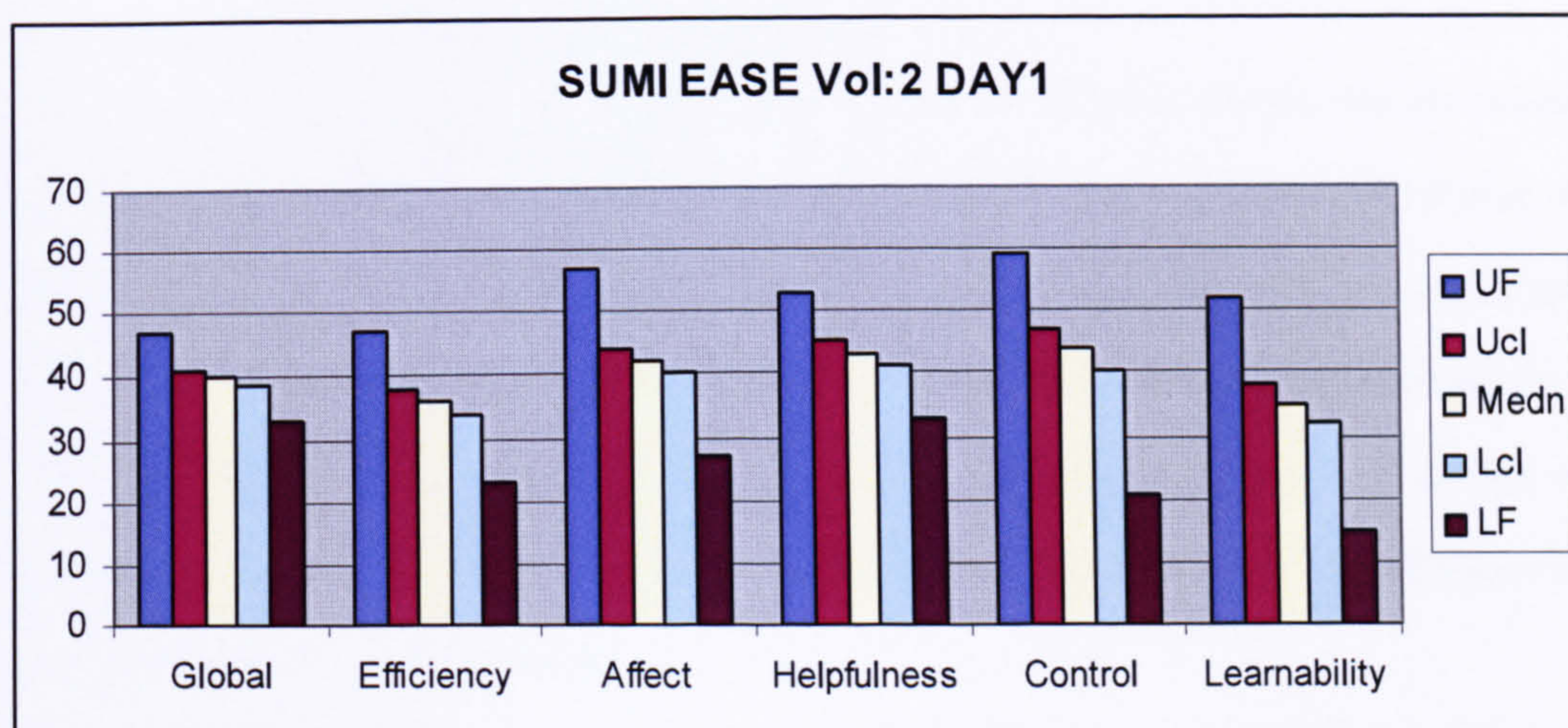


Figure 6.6 DAY-1 SUMI results of *Seminar Skills*

The difference in the two days results in the median score of the profile analyses in the results from two days is given below.

Table 6.4 Comparison of Day 1 and Day 2 SUMI results of *Seminar Skills*

	Day1	Day2
Scale	Medn	Medn
Global	40	63
Efficiency	36	63
Affect	42	60
Helpfulness	43	64
Control	44	57
Learnability	35	62

This concludes the section on SUMI DAY 1 and DAY 2 results. Informed by this analysis I will turn next to a discussion of questionnaires as an evaluation instrument.

6.4 Questionnaires as evaluation instruments

Questionnaires are a popular evaluation method because they are quick and efficient, but questionnaires have to be rigorously tested before they are used as an evaluation instrument (Nunan 1992; Cohen, Manion and Morrison 2003). In the interest of saving time and effort teacher evaluators can use pre existing questionnaires (such as PLUM), although they may have to adapt or change the wording of the items slightly to make the questionnaire more pertinent to what they are investigating. This needs to be done very carefully as sometimes an adaptation can lead to the question item being misunderstood or only partially understood, as in the case of PLUM Question 3b for Seminar Skills:

Could you please comment on any improvements you are aware of in: b) Your speech making skills?

Perhaps the question should have been worded differently as it did not capture the content of the program or confused the students about what was meant by speech making, as the answers to this question show ambivalence and confusion (Cf: Section 6.2.4.1). Questionnaires like SUMI are rigorously tested and their validity and reliability already established, so changing the wording of items may damage their rigour. Using them in their original form may result in fewer validity concerns.

Another issue regarding the use of questionnaires is about how carefully and earnestly the questions are answered. The following comment is a facetious or cynical one:

1c: Yeah, my typing speed improved which I was not expecting by this class. Q-49

The typing speed may have improved but it was certainly not being taught and it is odd for the student to mention it; this particular student has made similar comments in response to other questions.

There is the possibility that other students did not fill out the questionnaire very seriously for there are instances where some sub questions are left blank. On the other hand one student towards the end of the questionnaire gave this comment:

4: Yes, because it's interesting and going over it again would help us in our professional lives as well. Q-038

This was a surprising comment from this student as he had been sceptical about the program and in an earlier part of the questionnaire made comments like: *1a: Learned everything, honestly! 1b: Nothing seriously.* From this

example the tentative conclusion can be drawn that respondents may eventually give serious answers at some point during the questionnaire.

Another important point of discussion that has emerged from this study regards the timing of the evaluation. The fact that SUMI DAY 1 evaluation of EASE is significantly different from DAY 2 raises questions regarding the reliability of the instruments. This difference in DAY 1 and Day 2 results can be explained from the 'context of use' guidelines given in the SUMI handbook. The students were given the questionnaire on Day 1 at the end of a long session when they were tired. Day 2 evaluation was conducted a week later after the respondents had made 30-minute presentations and had the opportunity of putting the learning from the materials to actual use. The timing of the Day 2 evaluation also became significant because it was conducted in the morning (9.30 am) when students were fresh. This issue became evident because SUMI was repeated. However the same phenomenon could occur with other methods as well, not just with questionnaires. The timing of an evaluation could often have an impact on results from a number of methods.

6.4.1 Overall results from the perspective of an evaluation of EASE

Both questionnaires evaluate the same features of the software, but differ in scope, style and primary aim. The PLUM scales of Question 2 correspond to the SUMI scales of *Global, Efficiency, Affect, Helpfulness, Control, Learnability*.

Table 6.5 PLUM and SUMI's correspondence in scales

PLUM Scales	SUMI scales
Question 1:Hopes/Expectations	Global /Affect/Learnability
Question 2: Description of program	
V1: Easy to operate =	Usability
V2: Enjoyable to use =	Affective
V3: Provides good support for the exercise =	Helpfulness
V4: Provides good advice on how to approach the task=	Helpfulness
V5: Helps you learn =	Learnability
V6: Fits well with the rest of the course =	Global
V7: Well worth the time spent on it=	Affect /Global
Question 3: Improvements in knowledge skill etc.,	Global/Affect/Learnability
Question 4: Would you use it again.	Affect/ Learnability

The teacher evaluator of the EASE materials would get post use evidence on the aspects shown in Table 6.3. This table shows how the items in both questionnaires correspond with each other and aim to measure more or less the same features of the materials. Overall both the EASE CD-ROMs received a favourable rating, with some aspects of materials getting more approval than others. *Seminar Skills* was rated more highly on all scales as compared to *Listening to Lectures*, as there was a closer match between the perceived needs of the students and the content of the materials.

Please see further sections 6.2.3 and section 6.2.4 (for PLUM) and section 6.3.2 and section 6.3.3 (for SUMI) for a more detailed discussion of results.

6.5 PLUM and SUMI's performance on the evaluation criteria

The performance of the two questionnaires on the core criteria is rated as negative or positive and degree is shown by negative and positive symbols.

Rating Scale: Degree expressed by the number of positive or negative symbols
Negative: (-,-,-) very poor/ (-,-) poor/ (-) slightly poor
Positive: (+)ok/ (++) fair /+++good/++++very good

Table 6.6 Questionnaires' Performance on Criteria

Criteria	PLUM Questionnaire	SUMI Questionnaire
1. Cost Effectiveness		
Time	Inexpensive (+++)	Inexpensive (+++)
Effort	Inexpensive (++)	Slightly more effort required. (++)
Money	Inexpensive(++++)	Inexpensive (++++) for research free version
2. Ease of Use		
Preparation	Some preparation (++)	Considerable Training and Reading of SUMI handbook context of use required. (++)
Implementation	Fairly Easy (++++)	Get a licensed version for research purposes. (++)
Data preparation	Transcription/coding/quantitative analysis (-,-)	SUMISCO analysis (++++)
3. Bias		
Researcher	No Bias (+++)	No Bias (++++)
Respondent	Inevitable (-)	Meant to elicit opinions so bias inevitable (-)
4. Ecological Validity		
Researcher Intrusiveness	None but Possible (+++)	None unless as the administrator of questionnaire (+++)
Environment Intrusiveness	Yes conditions of use may influence (-,-)	Yes conditions of use may influence, if respondents are tired(-,-)
5. Nature of Information		
Depth	Yes(++) (In few cases respondents are not serious)	Yes (++) (In few cases respondents are not serious)
Objectivity	Fairly objective (+++)	Very objective (++++)
6. Immediacy of Response		
Immediate	No , Data has to be processed (-,-)	No, Data has to be processed (-,-)

Delayed	Yes (-,-)	Yes but not too much time used (-,-)
7.Usability measure		
Satisfaction	Yes (+++)	YES measures attitudes and satisfaction is designed to do so. (++++)
Effectiveness	Yes (+++)	YES especially designed to do so(++++)
Efficiency	None to average (\pm)	YES High as especially designed to do measure this (++++)
8.Robustness of Method		
Validity	High but has to be established (+++)	High already established (++++)
Reliability	High (in the same context) (+++)	High (In the same context) (+++)
Productivity	Average (+) The open ended questions did enable the method to elicit some new unpredicted before issues	High (++++).As predicted by the number of items on the inventory but fairly high

Both the *PLUM* and *SUMI* questionnaires are designed to test the learning effectiveness and usability of MM learning software. They were economical in time usage. PLUM required no effort to implement as it allowed for adaptation within its question design. SUMI on the other hand required more effort in terms of understanding the conditions of *context of use* stipulated in the accompanying handbook and acquainting oneself with the statistical background of the inventory before implementation. The researcher and teacher evaluator is required to understand the development and statistical standardisation background of SUMI. There was no financial layout for both questionnaires as PLUM was free to use from the Open University website and SUMI was licensed for free use as it was being used for research in an educational context.

On the criterion of *ease of use* considerable preparation was required for SUMI prior to implementation and context of use requirements had to be met. Implementation of both questionnaires was easy once the SUMI version licensed for research was received. Data preparation for PLUM was more time and effort intensive as data had to be analysed both quantitatively and qualitatively whereas SUMI was easy and convenient as the especially designed analysis tool SUMISCO conducted the analysis.

No researcher *bias* was evidenced in this method, since the method was designed to capture respondents' attitudes and preferences and therefore bias was not applicable to respondent comments.

There was no direct *researcher interference* evidenced in the method: the administrator of the questionnaire was a teaching assistant who was instructed to remain neutral and merely distribute the questionnaires. Environment and context *intrusiveness* however may have impinged on the data and it could have been influenced by the timing of the evaluation. At the end of a long tiring day if the respondents are asked to fill in a 50 item questionnaire

they may not be very positive or honest. This is indicated by the two SUMI evaluations. The first time the evaluation was conducted it was at the end of a long evening session where students had just finished working on the materials. The second time it was conducted at the beginning of a debriefing session, after the students had time to reflect on their learning, and a substantial amount of assessment for the course had already taken place. SUMI evaluations can be done any number of times at various stages of working through the software. So this repeat evaluation was not a departure from accepted practice and it yielded results which had serious implications for the method's reliability. The timing of the evaluation becomes an important factor here.

On the criterion of *nature of information* both PLUM and SUMI revealed fairly deep data as they were designed to question attitudes and preferences in relation to the materials. However, in certain cases where the students filled the questionnaires non-seriously, the information was superficial, particularly in the open ended questions of the PLUM questionnaire. The design of question items in SUMI and its rigorous reiterations pre-empted this and there was no scope for any surface level comments to be recorded by the respondents.

In both questionnaires the data has to be processed and analysed, which may take some time, so on the scale of *immediacy of response* both PLUM and SUMI do not score highly and results and impressions are delayed. The SUMISCO analysis tool was more efficient but access to the analyst is required, or alternatively skill in using the tool has to be acquired.

Both the questionnaires measured *usability* aspects effectively as both questionnaires were designed to evaluate usability. PLUM measured *satisfaction* to an average extent whereas SUMI does this more effectively as the questionnaire is specifically designed to measure usability. SUMI scores high on the scales of *effectiveness* and *efficiency* as well. The PLUM questionnaire scored low on these scales as compared to SUMI but on its own it delivers on this criterion of obtaining valuable data on usability issues.

On the scale of *robustness of method* questionnaires scored highly because the *validity* of the method is upheld: it measures what it is designed to do. The *reliability* of the method is also high. Provided it is used in the same context (without any issues like timing impinging on the mood of the respondents) it will yield replicable results. The *productivity* of the method is catered for in the design of the question statements, which have probably evolved from usability heuristics and guidelines for effective design. The productivity is high because it yields

more issues than the other methods and even certain additional dimensions are revealed which may not have been part of the original intention of the questionnaire design.

6.6 Summary

This chapter has reported on results from two kinds of questionnaires used to evaluate EASE materials. To understand how these questionnaires performed as evaluation instruments, elicited data was first analysed and then reported as an actual evaluation. Studying these methods in use enabled me to evaluate the method in relation to the evaluative criteria at the heart of this study.

This chapter dealt with the objective quantitative method of questionnaires. I now turn to another objective method of evaluation which turns the focus on what learners actually did (as opposed to what they tell us they did) as they work on the materials. Tracking data about user behaviour collected through activity monitoring and key logging software is discussed next.

CHAPTER SEVEN: ACTIVITY MONITORING AS AN EVALUATION METHOD

7.1 Introduction

Chapter two (section 2.6.4) gave a general description of this method. Chapter three (section 3.6.4) gave the rationale for using it and describes the software involved and its functionality. It also explains the procedure for using the activity monitoring software in the main study, which was informed by the findings of the pilot study.

This chapter first explains the data analysis procedures that were adopted for this study. The second section gives an account of applying these procedures to *EASE* Volume 1: *Listening to Lectures*, beginning with flowchart representation of the students working through certain units and then discussing the learners' navigation patterns and time usage through these units. The third section replicates the second section, but with reference to *EASE* volume 2 *Seminar Skills 1: Presentations*. The performance of the method on the evaluation criteria is discussed in the fourth section, first from the perspective of an evaluation of *EASE*, and secondly from the perspective of activity monitoring as an evaluation method.

7.2 Data reduction and analysis procedures

Observation using tracking and activity monitoring software, by its very nature is a close and rigorous method of evaluation. I gathered activity monitoring and key logger data from 40 plus students in the main study and was obliged to undertake a random selection of data sources (Cf Chapter 3 Section 3.6.4). A further selection had to be made when it came to presenting findings in this chapter. Further information is provided in Appendix 7. Although data from activity monitoring can be statistically analysed through an especially designed computer application to make the interpretation of the data easier (as done by Hwu (2003) see section 2.4.2.5), I decided to analyse the reduced data more qualitatively by studying the navigation patterns. This was seen as being in keeping with the constructivist case study approach of this study.

The exploration of the data was guided by the following questions, which I felt that a teacher evaluator would be likely to ask:

1. *What is the level of engagement with the software?*
2. *How much time is spent on each exercise/page?*
3. *What is the average speed of completion of a unit when done properly?*
4. *How often are the additional features of the program (e.g. dictionary) accessed?*
5. *Are the students distracted, and if so, how?*
6. *How does this distraction impact on their performance?*
7. *Overall what are the learning implications from all these observations?*

After retrieving the data the next stage was to put it in a comprehensible format. Microsoft Visio was used to create detailed flowcharts of the navigation of the CD-ROMs. Since this was a time consuming exercise and the data available was copious it was decided that limiting the number of students observed and selecting two units from each CD-ROM could be a manageable strategy. Unit 2 and 6 were selected from *EASE Volume 1 Listening to Lectures* and Unit 1 and 5 from *EASE Volume 2 Seminar Skills*. The selection of two units from each CD-ROM was pragmatic rather than random. The intention was to study a unit from the beginning and a unit from the end of each CD-ROM, but the activity monitor was not fully functional in the first session where some students started with Unit 1 of Volume 1 *Listening to Lectures*. Five students' records were randomly selected for each unit. The computer (Microsoft Excel RAND function =RAND ()*70) was used to generate a random selection. The flowcharts were created by first creating a generic flowchart of the unit and then charting the course of each student's navigation, represented by red coloured lines. Six flowcharts were created for each unit: one generic flowchart and five showing each of the five selected students' navigations.

7.2.1 *Navigation patterns*

The navigation patterns were first determined by mapping the individual student's course through the CD-ROMs on to the generic flowchart of each unit. Then because of the difficulty of representing the entire flowchart in compact form, the patterns of each of the selected students' navigation were collated in a picture diagram as shown in Figures 7.7, 7.10, 7.17, 7.19.

7.2.2 Time usage

Studies have established a correlation between the amount of time spent on a page of materials with learning effectiveness (Hwu 2003; Quentin-Baxter, 1999). Statistical studies by Hammond and Allinson (1989) and Quentin-Baxter (1999) support the correlation between amounts of time spent interacting with materials and effectiveness of learning. Hammond and Allinson's (1989) study upholds Baddeley's (1976) *total time* hypothesis which states that the amount of time spent learning a material is directly proportional to the amount of learning. Quentin-Baxter (1999), reports that students who did not access much information because they did not spend enough time working on the materials had a poorer and less accurate appreciation of their own achievement.

My aim was to observe the method at work and I took the view that observation through activity monitoring which captures the learner at work through screen shots was a richer medium of observation than just collecting audit trail data like tracking or key logger data.

The key logger data, though useful for counter checking, did not provide the level of detail required to determine what the students were actually doing. Studying the screen shots and noting the amount of time spent on each page and exercise made for greater clarity and understanding of what was happening in the recordings (Cf. Appendix 7.1). The Table 7.1 below is an excerpt from key logs of two students' work showing their typing errors or spelling mistakes. This table puts keylogger data from two students PC53 and PC54 side by side. The information has the time and date the name of the program, the unit the students are working on (both are working on Unit 2 working on a lecture on 'cinema'. The information which follows <<SCR>> gives the screenshot picture number which has an ending of 'jpg.'

Table 7.1 Key logger reports from two students

Key Logger data from Student PC53 (Excerpt)	Key Logger data from Student PC54 (Excerpt)
<<PROC>> <<TIME 12/20/04 17:45:18>> C:\Program Files\EASE\EASE1.exe <<WND>> EASE: Unit 2/Structure and organisation <<SCR>> <<TIME 12/20/04 17:45:26>>	<<BEG>> <<TIME 12/15/04 18:41:56>> Logging started on <<USER 08020091>> <<COMPUTER LAB2PC54>> <<SCR>> <<TIME 12/15/04 18:41:57>> 041215_184157.jpg

<p>041220_174526.jpg</p> <p><<SCR>> <<TIME 12/20/04 17:45:41>> 041220_174541.jpg</p> <p><<SCR>> <<TIME 12/20/04 17:45:56>> 041220_174556.jpg</p> <p><<SCR>> <<TIME 12/20/04 17:46:11>> 041220_174611.jpg</p> <p>Italian Cinema 19</p> <p><<SCR>> <<TIME 12/20/04 17:46:26>> 041220_174626.jpg</p> <p>45-790</p> <p><<SCR>> <<TIME 12/20/04 17:46:41>> 041220_174641.jpg</p> <p>Course: Title</p> <p><<SCR>> <<TIME 12/20/04 17:46:56>> 041220_174656.jpg</p> <p>Lecture : Introductio</p> <p><<SCR>> <<TIME 12/20/04 17:47:11>> 041220_174711.jpg</p> <p>n to Neo-Realism</p>	<p>people 2nd: Popular Genre Cinema: For the Pei</p> <p><<SCR>> <<TIME 12/15/04 18:42:13>> 041215_184213.jpg</p> <p>people 3rd Part: Work of Fe</p> <p><<SCR>> <<TIME 12/15/04 18:42:27>> 041215_184227.jpg</p> <p>dric Feline: Work hjaad</p> <p><<SCR>> <<TIME 12/15/04 18:42:42>> 041215_184242.jpg</p> <p>neo-realism pictures but also popular gernre</p> <p><<SCR>> <<TIME 12/15/04 18:42:57>> 041215_184258.jpg</p> <p>had works in cinema. in beginin</p> <p><<SCR>> <<TIME 12/15/04 18:43:12>> 041215_184312.jpg</p> <p>nigng of career. Was Ambigious.</p> <p><<SCR>> <<TIME 12/15/04 18:43:27>> 041215_184327.jpg</p> <p>Main Pioints: tTrying to capture</p> <p><<SCR>> <<TIME 12/15/04 18:43:42>> 041215_184342.jpg</p> <p>spiirit of Italian cinema Workl</p> <p><<SCR>> <<TIME 12/15/04 18:43:57>> 041215_184358.jpg</p> <p>ds for thethat were poplu</p> <p><<SCR>> <<TIME 12/15/04 18:44:12>> 041215_184412.jpg</p> <p>ular withth the classes People wanted to see epic ands</p> <p><<SCR>> <<TIME 12/15/04 18:44:27>> 041215_184427.jpg</p> <p>comsedys Course Outlienne:</p> <p><<SCR>> <<TIME 12/15/04 18:44:42>> 041215_184442.jpg</p>
---	--

	<p>Readings: Good starting points, not all the material</p> <p><<SCR>> <<TIME 12/15/04 18:44:57>> 041215_184457.jpg</p> <p>required to be read , D</p> <p><<SCR>> <<TIME 12/15/04 18:45:12>> 041215_184512.jpg</p> <p>istcriminate between the texts.</p> <p><<SCR>> <<TIME 12/15/04 18:45:27>> 041215_184527.jp</p>
--	--

7.2.3 Narratives of navigation

A third attempt at data reduction was undertaken by writing a narrative account of what the students did while working on each unit of the materials. This exercise gave greater clarity to the flowcharts and their later interpretations.

7.3 EASE Listening to Lectures

A description of the materials is given in Chapter One section 1.2.3. A structural chart with pictures of screenshots of the *EASE Volume One Listening to Lectures* is given in Appendix 1.1. (EASE CD-ROM included as Appendix 2.1). A description of the units selected (2 and 6) for closer scrutiny is given in Appendix 7.2

7.3.1 Flowchart of Listening to Lectures Unit 2

Figures 7.1, 7.3, 7.4 and 7.5 below show graphic representations of MS Visio flowchart compressed to a degree which makes them presentable here. The generic flowcharts show the complete layout of the unit. The top of the flowchart (fig 7.1) with seven 'diamond shapes' give the six units which the user can select. On account of the degree of compression it is not possible to read the writing inside the shapes. Figure 7.2 shows the detail of one section of the flowchart and the captions in the shapes are visible. The diamond shapes in the flowcharts show stages or learning units where an option can be exerted by the user it is a 'decision diamond'. The rectangle shapes give options which are common to all units (Cf: Appendix 7 and the accompanying appendix 8 CD-ROM for full VISIO flowcharts).

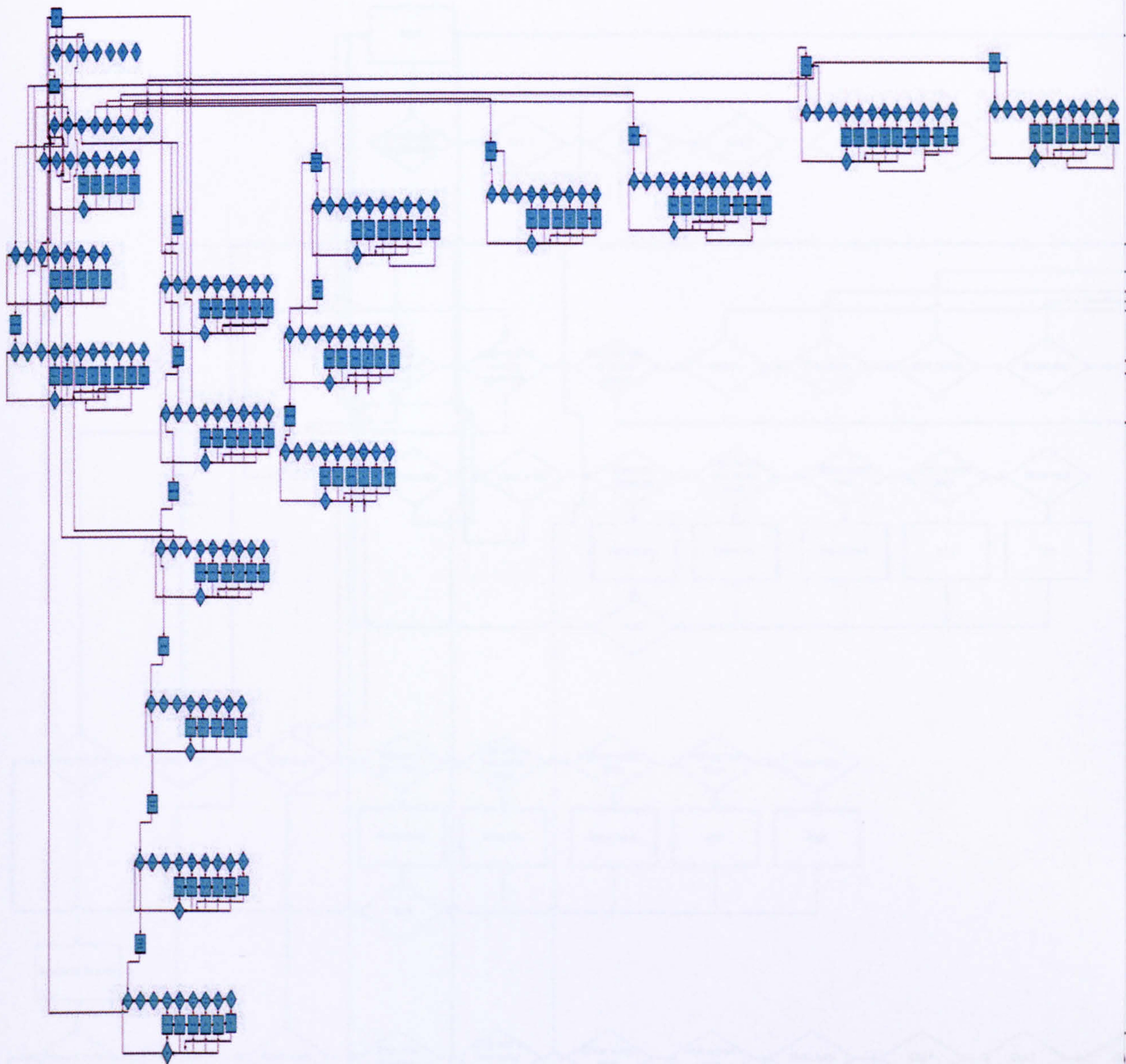


Figure 7.1 Generic Flowchart of EASE Vol 1: Listening to Lectures Unit 2

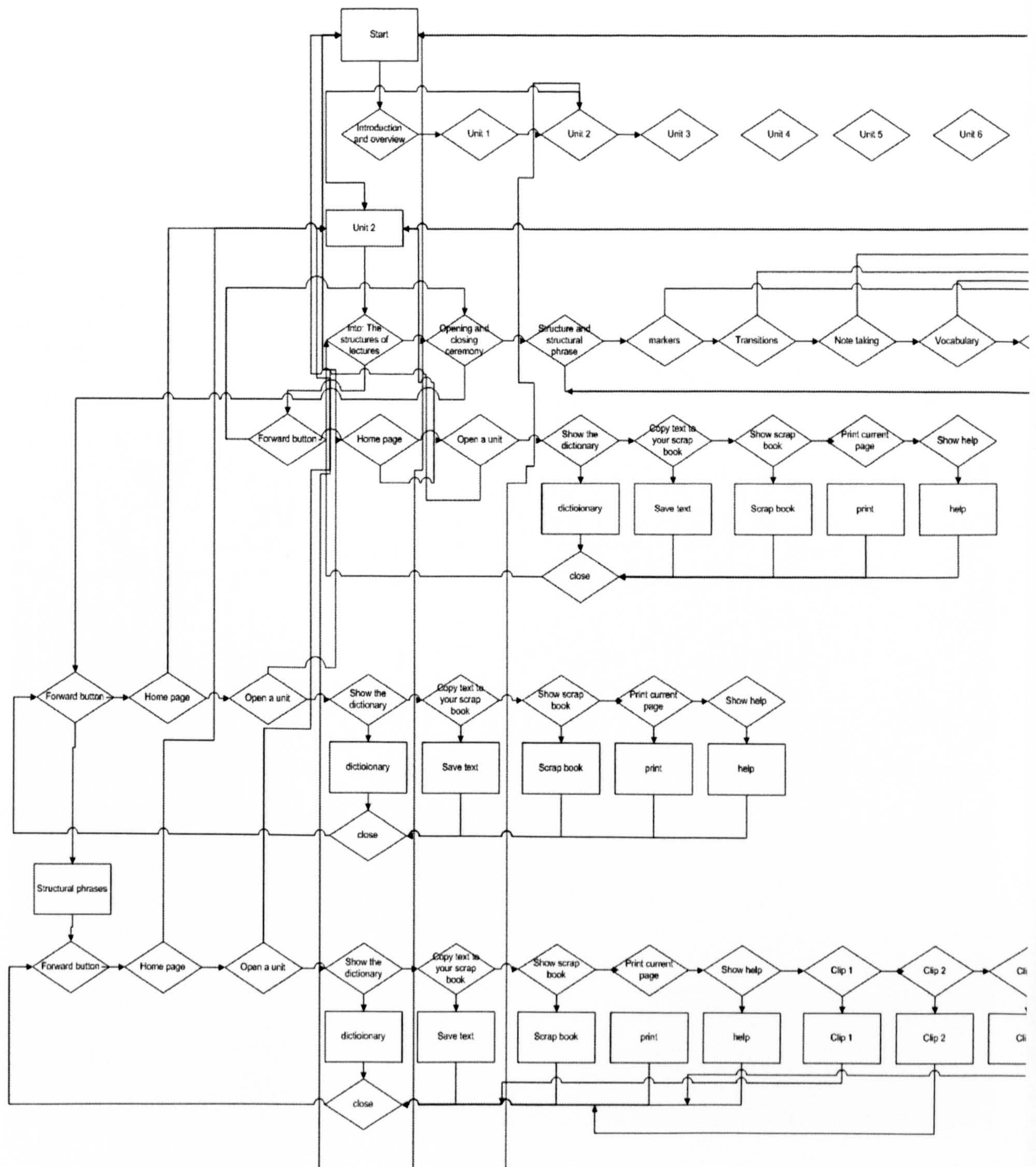


Figure 7.2 Closer look at start of unit 2 Vol1

7.3.2 Flowchart of Listening to Lectures Unit 6

This flowchart gives the complete generic structure of Unit 6.

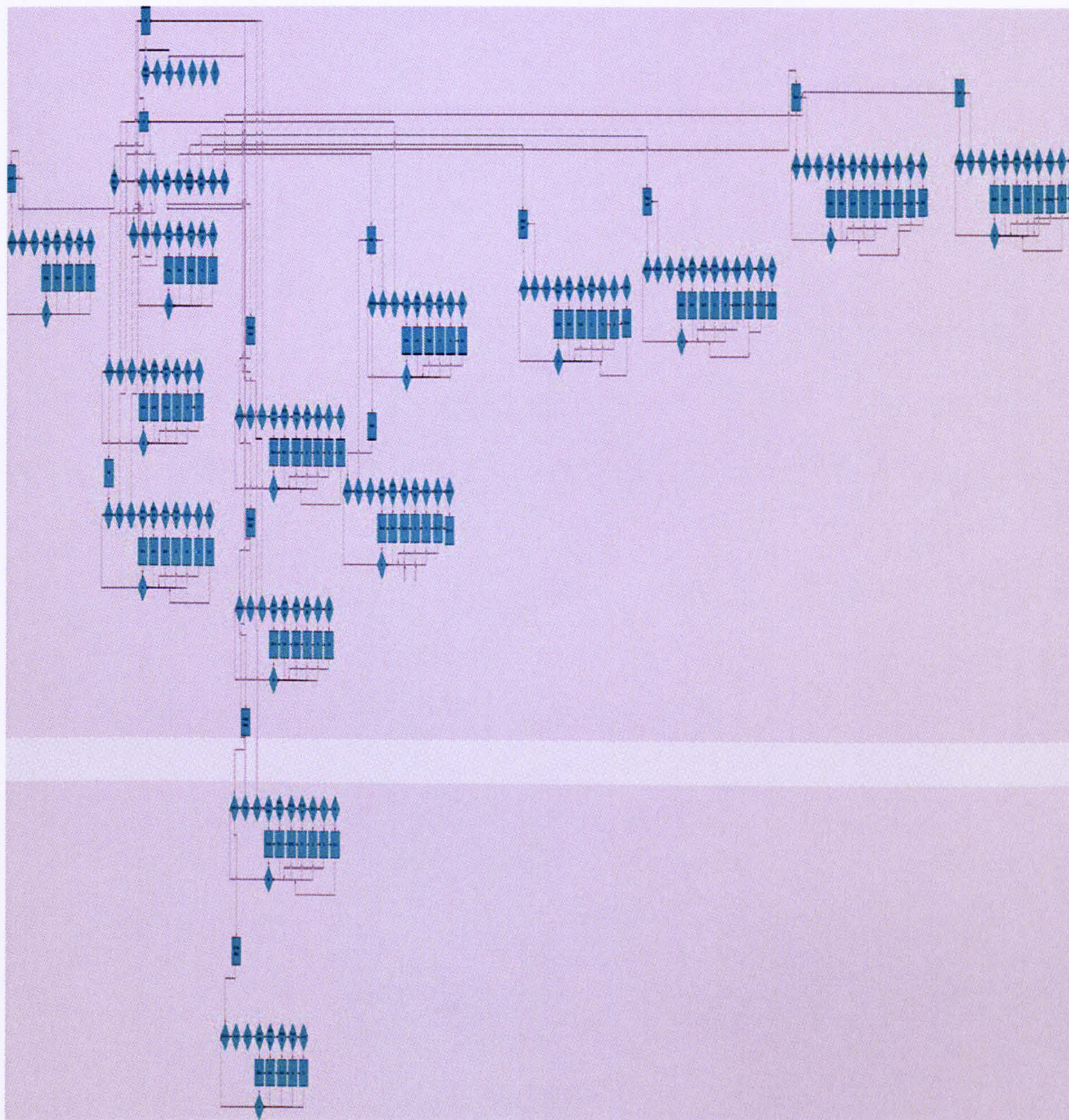


Figure 7.3 Generic Flowchart for Unit 6 of Listening to Lectures

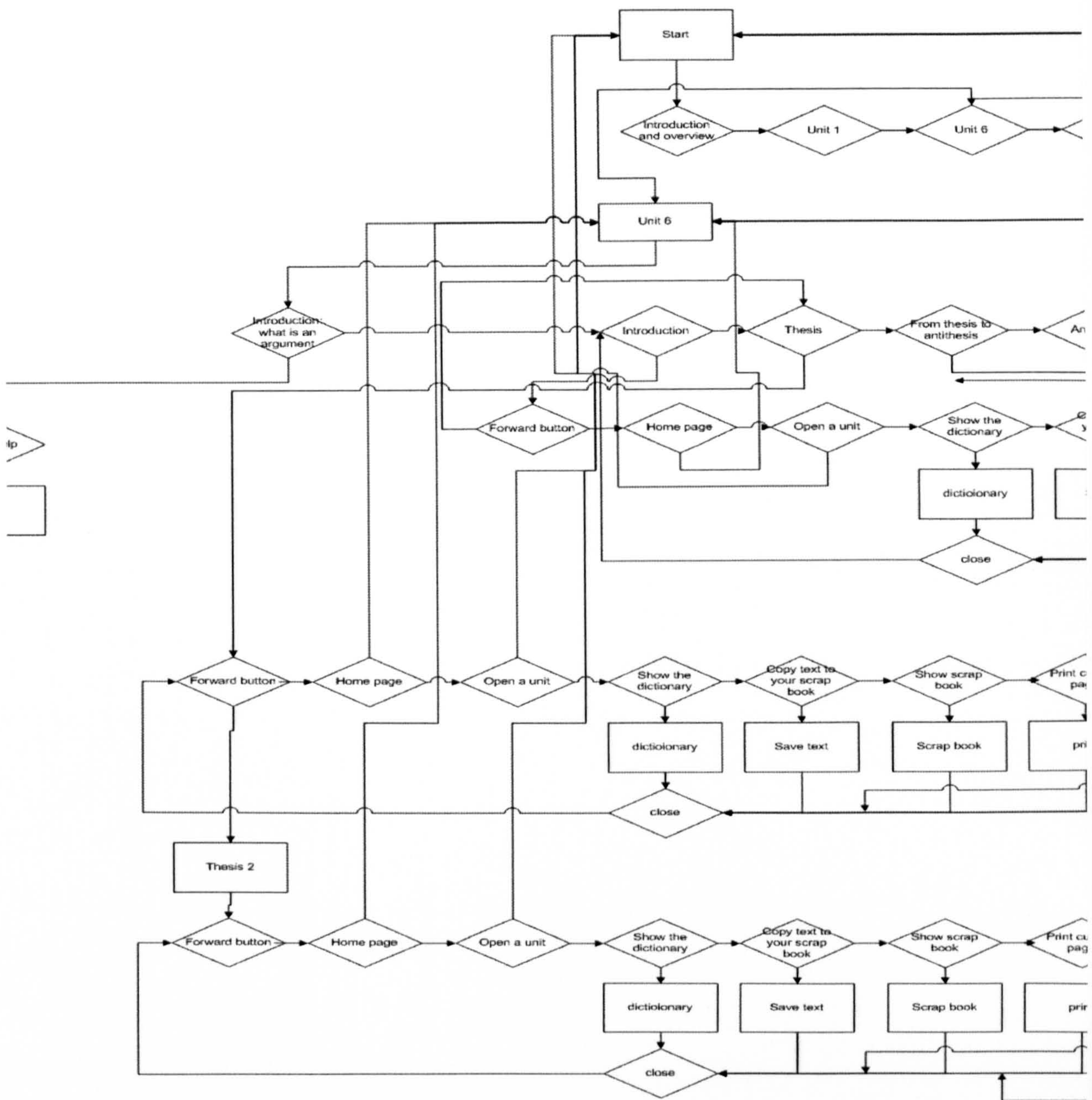


Figure 7.4 Closer look at Unit 6 Vol 1

7.3.3 *EASE Listening to Lectures Unit 2: Structures and Organizations*

For a complete explanation of the chosen units from *EASE* volume 1 *Listening to Lectures* see Appendix 7.2

In Unit 2 *EASE Listening to Lectures* there are a further 8 modules. For complete descriptions of Unit 2 see Appendix 7.2

1. Introduction: The structures of lectures
2. Opening and closing summaries
3. Structure and Structural phrases
4. Markers
5. Transitions
6. Note taking
7. Vocabulary
8. Conclusion

7.3.3.1 *EASE Listening to Lectures: Flowcharts and Narrative of Navigation*

Figure 7.5 shows PC 53's working through Unit 2. The picture quality is poor due to compression and the captions within shapes are not visible. The 'red coloured' lines depict the navigation route of PC53. The red lines were drawn on the generic flowchart of Unit 2 (Fig.7.1/7.2). This flowchart picture has been cropped from the right and bottom but the student PC53 had not extended that far right nor below.

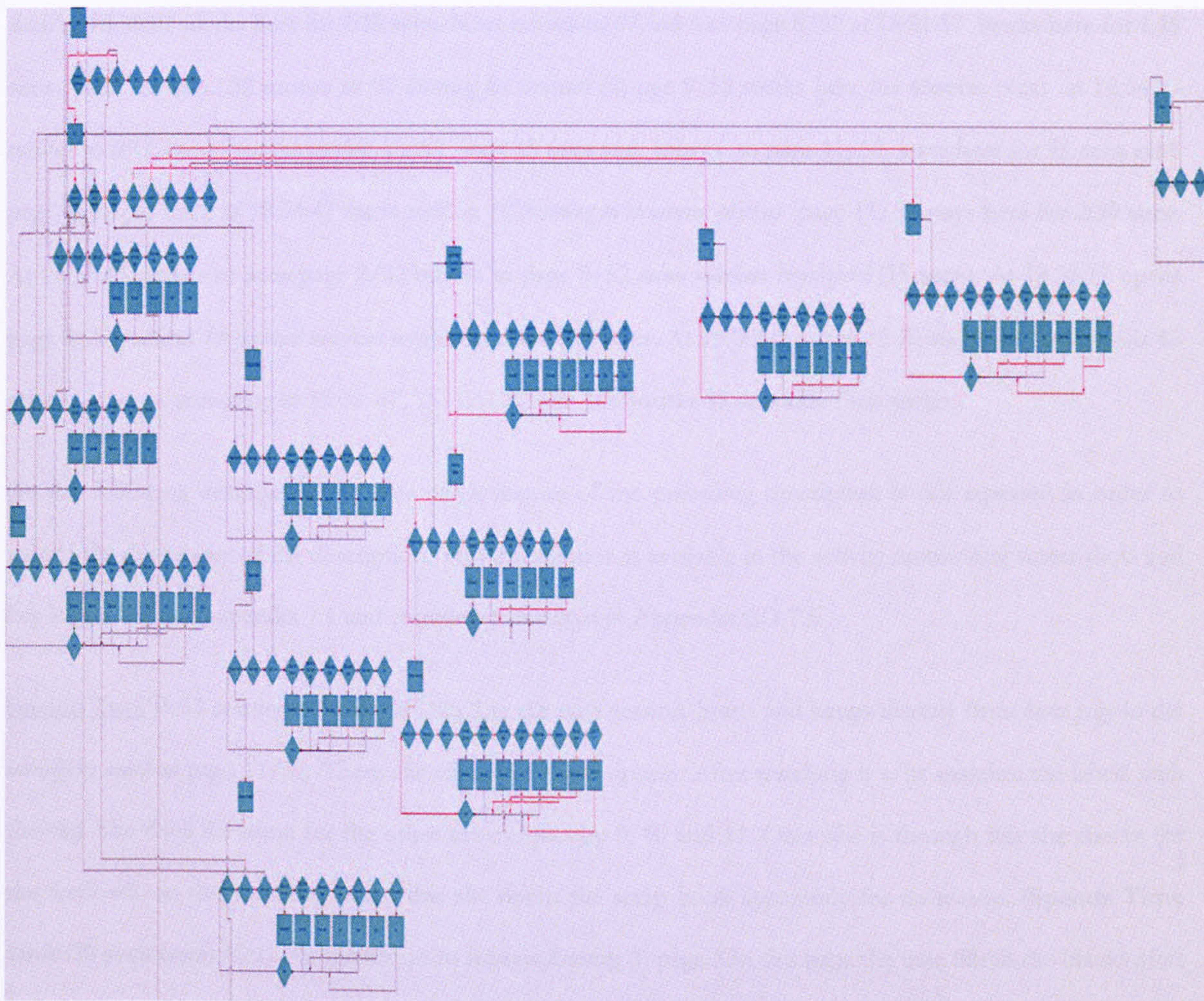


Figure 7.5: PC53 (session 1 on Unit2) Navigation Pattern shown by red coloured lines.

Narrative accounts were written by studying the flowcharts of five students' work. Only one student's account is presented here in detail in order to illustrate the possibilities of data interpretation from this method.

Narrative of Navigation of PC 53

Session One: Screenshots were recorded at 15 seconds intervals. The student started work on this unit at 18:40:17 opens page 1/52 Structures and Organisation of lectures spends **1 minute** reading instructions. 18:41:17 Goes to 01 *lecture summaries* (i) page 2/52 spends **2:45 secs** working on this section. 18:44: 02 goes to 02 section *Lecture Summaries* (ii) stays here **30 secs**. Opens page 03 *Listening for structure* at 18:44: 32 page 4/52 works here for **1:45 secs**. Next moves to 04 *structural phrases* page 5/52 at 18:46:17 works here for **2:45secs**. Next moves to 05 *Subject matter* page 6/52 at 18:49:17 stays here for **30 secs**. Moves to 7/52 sec 06 *Listening for*

detail at 18:50:02 works here for **1:15 secs**. Next moves to *07 verb tense* page 8/52 at 18:51:17 works here for **1.15 secs**. Next at 18:52:32 moves to *08 listening for structure (ii)* pge 9/52 works here for **45secs**. Next at 18:54:17 moves to *09 Listening for structure(iii)* 10/52 stays **15 secs** and moves to page 11/52 stays here for **15 secs** each page perhaps. Then at 18:54:47 starts section 10 *listening to structural phrases* page 11/52 stays here for **2:30 secs**. At 18:57:47 moves to *menu* page 2/52 moves to page 8/52 *menu markers highlighted (15 secs)*. At 18:58:17 opens page 9/52 section 11 *forward markers* works on it for **1.30 secs**. At 19:00:02 opens 12 *Backward markers* spends **45 secs** here stops recording at 19:00:47. **TOTAL TIME 16minutes 15 seconds** First session.

(In the following description the time clock feature of the preceding description is not repeated in order to reduce the granularity of the description. This information is available in the activity monitoring screenshots and key logger data in Appendix 7.1 and complete Flowchart in Appendix CD 7.5.

Session Two: Pc53 continues work on Unit 2 in the next session. Starts and jumps directly from *home page* to the *identifying markers* page 11/52. There she clicks the clip 8 option. After watching it s/he matches the labels with the clip. She does the same for the other three clips, clip 9, 10 and 11. Once she is through this she checks for the feedback on the matching. After this she opens the scrap book apparently for no reason. (Spends Time **4min:30 secs** here) Next she moves on to *intensive listening (i)* page. On this page the user fills in the blanks after seeing the clip 8 again. Then checks for *feedback*. After this she refills the blanks and checks for the answer and this time gets it right. The user does the same for the clip 9, 10 and 11 except for the feedbacks does not check them. Does *Intensive listening (ii) (iii) and (iv)* **Total Time on Intensive Listening 2:30+1min+45 secs+1:15 = 5:30 secs**. Next she moves to the *transitions* page. Over here she watches clip 12 and drags the labels accordingly and checks for the feedback. For clips 13 and 14 she does the same for what she had done for clips 8, 9, 10 and 11. **Total time transitions (i) (ii) and (iii) 3 :30**. Next she moves to the *rhetorical questions* page. Here the user sees the clip 15 and fills in the blank (Time **45 secs** only) Next she moves to the *Linking Exercise* page where the user writes on the fill in the blank space provided. Next she opens clip 16 and compares what she has written with the clip (*Linking (i) (ii) and (iii)* **6:45**). Next are the *identifying transitions* pages (22-24/52) where the user watches the clip. On the next page *identifying transitions (ii)* the user watches the clip 19 to 22 and decides the transitions, total time spent: **6 minutes**. Next is the *note taking* section 27 (i) 28 (ii) where the user makes notes

for 2 minutes. The user spends 29 minutes from *Identifying markers* to here. Total time of the two sessions on Unit 2 is 47minutes 25 seconds.

The above narrative account has been created by studying the screenshot and key logger data which were also consulted to create the flowcharts. Figure 7.7 is a screenshot from PC53 working on Unit 2.

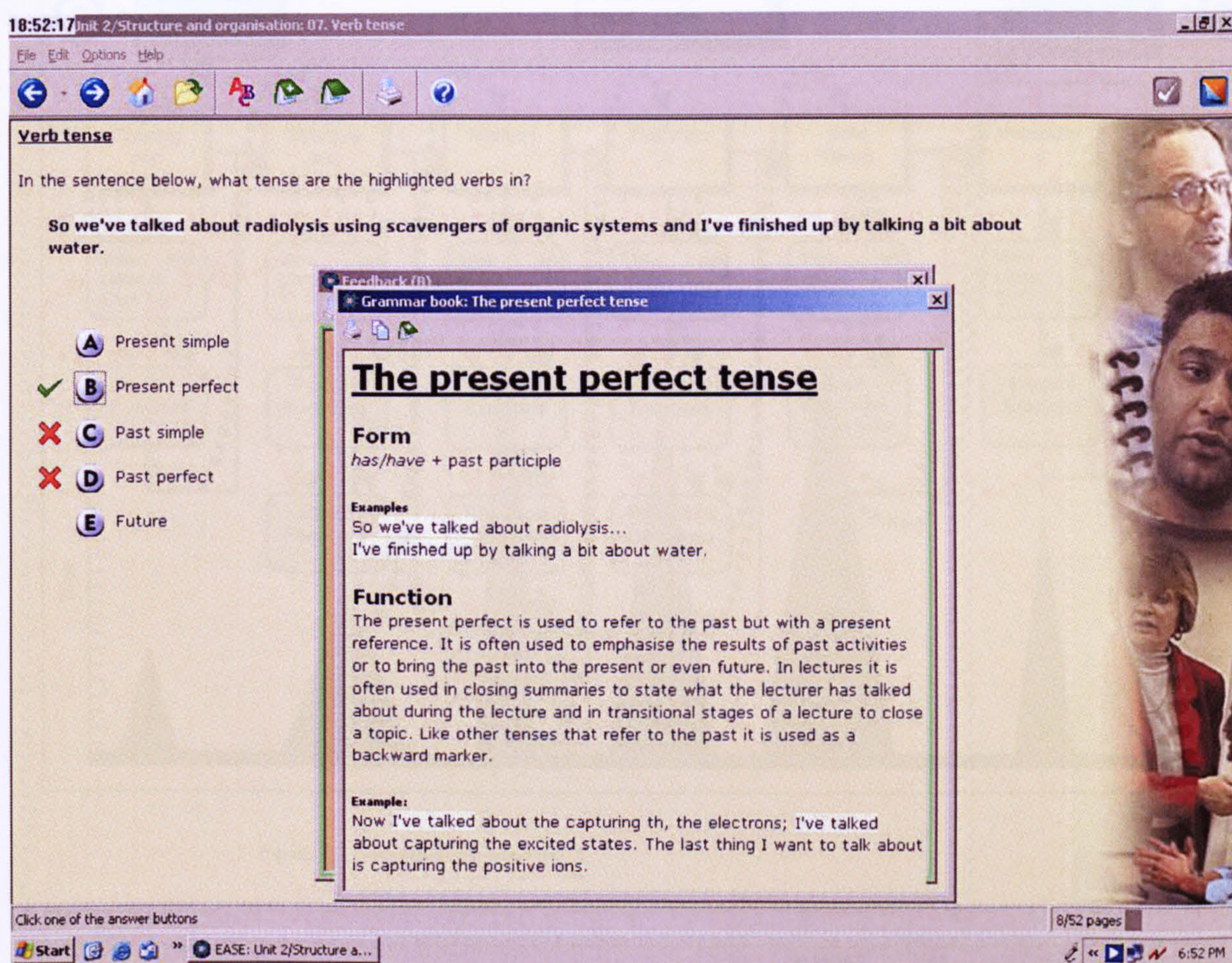


Figure 7.6 Student PC53 consulting the Grammar book

7.3.3.2 Five students' navigation patterns for Unit 2

The five students working through unit 2 are shown graphically in Figures 7.7 and Figure 7.8.

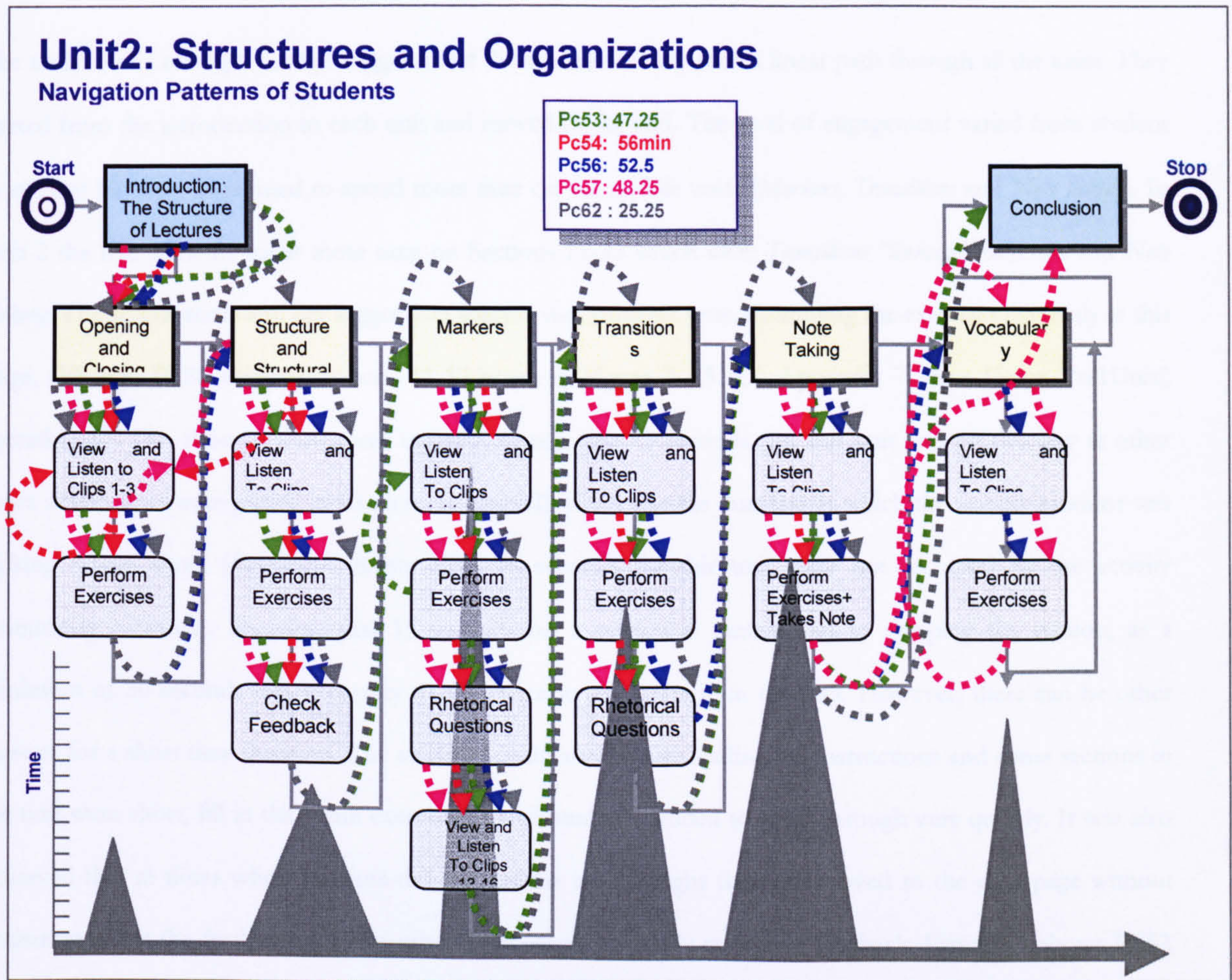


Figure 7.7 Navigation pattern Unit 2

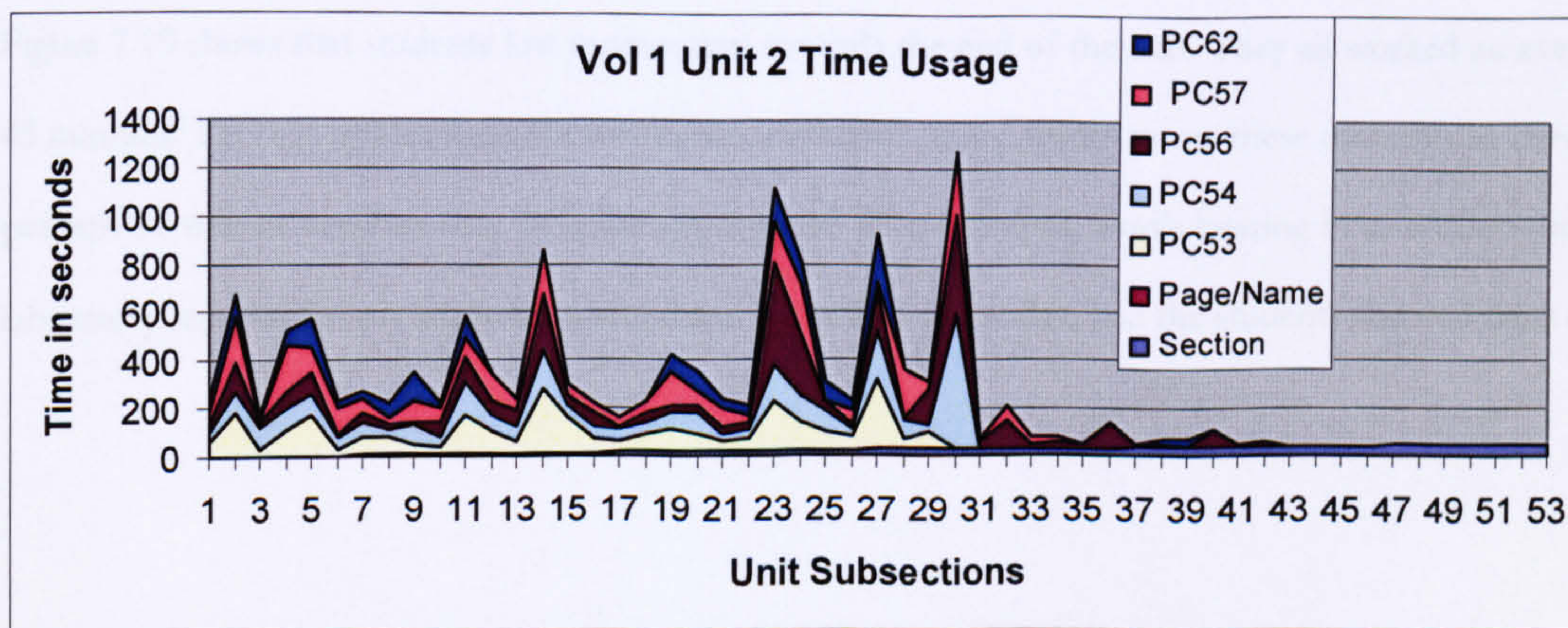


Figure 7.8 Time Spread Unit 2 Listening to Lectures

7.3.3.3 Analysis of five students working on Unit 2: Listening to Lectures

The analysis of *Listening to Lectures* suggests that the students mostly took a linear path through all the units. They started from the introduction to each unit and moved to the end. The level of engagement varied from student to student but they all seemed to spend more time on the middle units (*Markers, Transitions* and *Note taking*). In Unit 2 the five students spent more time on Sections 21-33 which were *Transitions "linking"* exercises and *Note Taking*. The screen shots and key logger data suggest that students were attempting the exercises diligently at this stage, although PC62 skipped sections 11-17 graph in Figure 7. 15 (Cf: Appendix 7 Time Usage Vol1Unit2 Spreadsheet). The shortest time spent on each section was 15 seconds (on this unit and on this day at other times screenshots were picked at 10 seconds interval) which was the duration at which the activity monitor was picking screen shots (durations shorter than 15 seconds for this unit were not recorded by the activity monitoring software). Spending just 15 seconds on a section is tantamount to skipping the section, as a minimum of 30 seconds was necessary to do justice to any exercise in the unit. However, there can be other reasons for a short time duration. The students could also be just reading the instructions and some sections in the unit were short, fill in the blank exercises, which students tended to work through very quickly. It was also observed that at times when students did not get the answer right they still moved to the next page without bothering to get the feedback. On the other hand others diligently consulted feedback. Figure 7.7 shows PC53 consulting the Grammar Book.

Four out of five students all spent more time on *Identifying Markers* section with only PC62 skipping this section. Figure 7.15 shows that students lost momentum towards the end of the unit. They all worked an average of 40-45 minutes. This has implications for the teacher evaluator. If she wants to use these materials in class time then perhaps 30 minute sessions may be more appropriate. It is, however, worth bearing in mind that the computer laboratory sessions for my study were scheduled at the end of the day, and the students showed signs of fatigue.

7.3.4 *EASE Listening to Lectures Unit 6: Argumentation*

In Unit 6 there are a further 10 modules:

1. Introduction: What is an argument?
2. Thesis: The lecture as an argument
3. From thesis to antithesis
4. Antithesis
5. The language and structure of an argument
6. Towards synthesis
7. A final question: Who produces the synthesis?
8. Note taking
9. Vocabulary
10. Conclusion

A complete description of Unit 6 in narrative form is given in Appendix 7.2

7.3.4.1 Flowchart of one student (pc 03) working on Unit 6

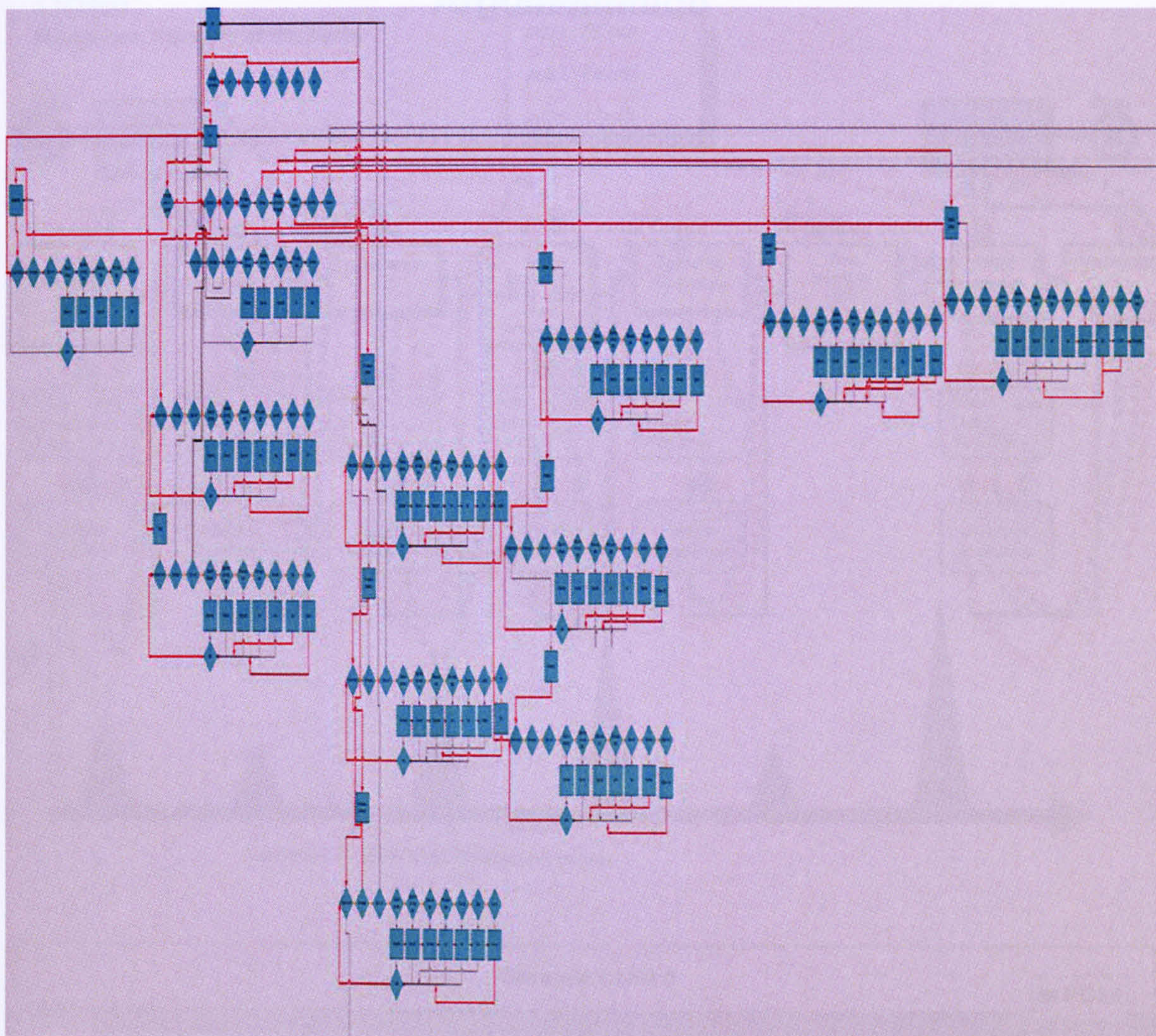


Figure 7.9 Flowchart of one student's navigation PC 03

7.3.5 Five Students navigation through Unit 6 of EASE Listening to Lectures

The following two figures show another set of five students work through Unit 6. The first figure (7.10) gives the navigation and Fig 7.11 gives the time spread across the worked unit. It also shows the names of exercises in the unit and the time spent on each exercise by these students is evident.

Navigation Patterns of Students

Argumentation

pc03 : 35 min
pc19 : 35 min
pc21 : 64 min
pc20 : 62 min
pc24 : 11min

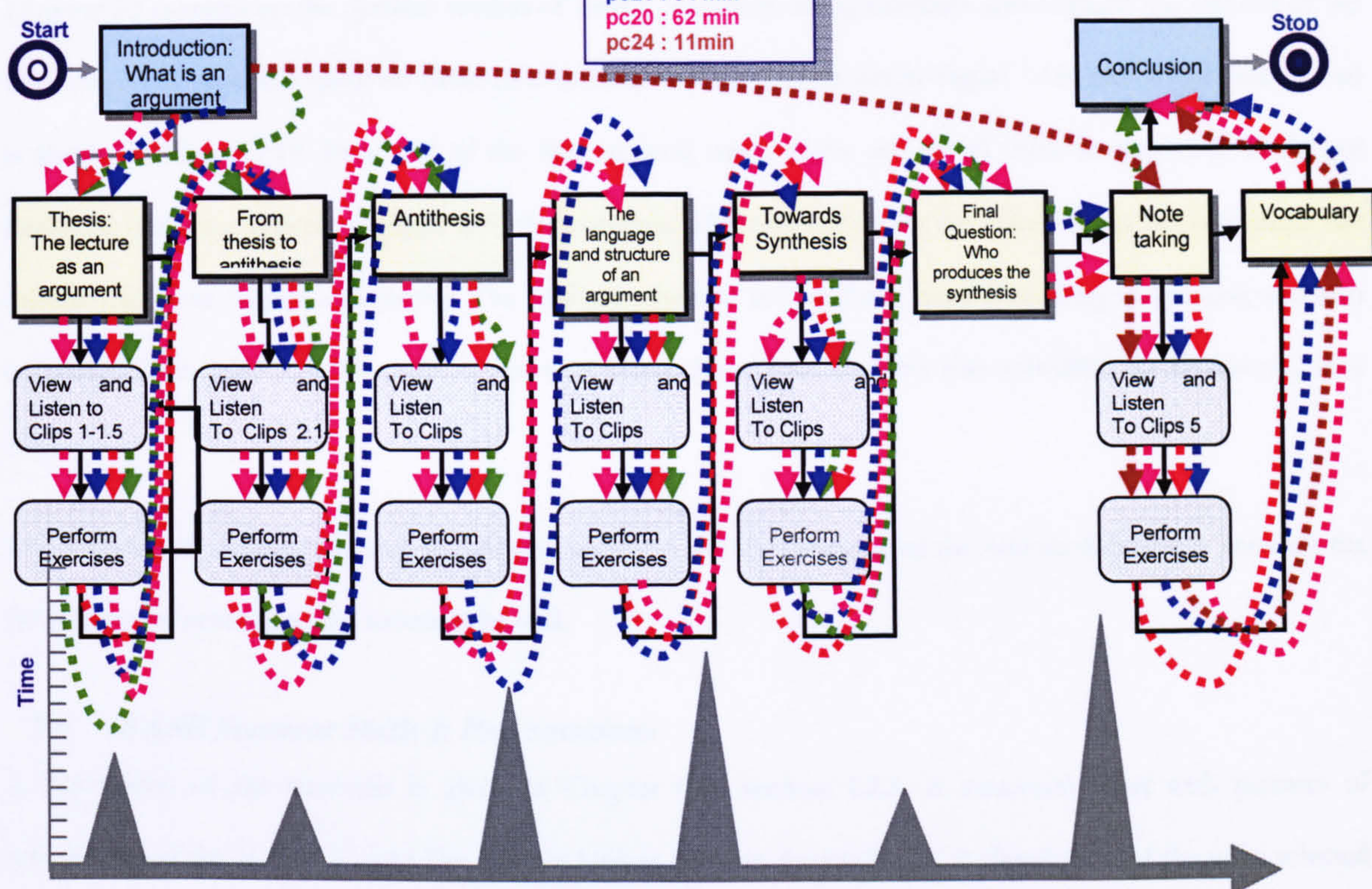


Figure 7.10 Unit 6 Vol1 Navigation picture

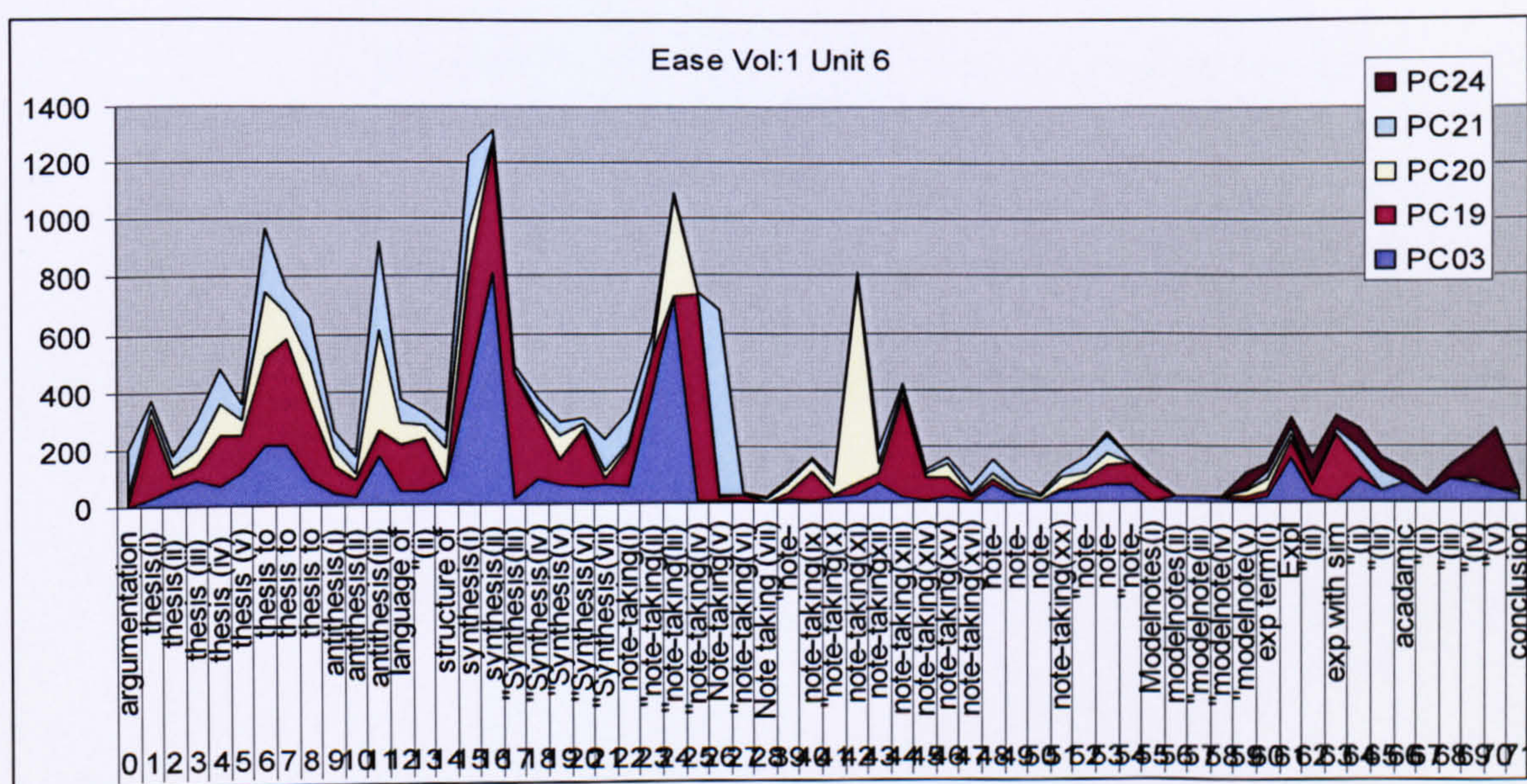


Figure 7.11 Unit 6 Vol 1 Time Usage

7.3.5.1 *Analysis of five students working on Listening to Lectures Unit 6*

The sections of the unit where students spent most time were “*Thesis*”, “*Antithesis*” and “*Synthesis*”. Student PC 19 spent 25 minutes on the *Synthesis* section of Unit 6. The note-taking exercises also engaged the interest of the students. Most time was spent on these four sections of Unit 6 as shown in Figure 7.18. The actual time spread is shown in Figure 7.19. Four out of the five students made notes and saved them in the scrap book, got feedback, and were actively engaged with the materials. This was evident in the screen shots but is perhaps not captured in these figures and graphs. The students checked for feedback quite frequently in this unit and also compared their notes with the authors’ notes to assess them. Two students also consulted the dictionary, once each.

The students all stopped working at different points in the later sections of the unit so differences between the flowcharts are most apparent towards the end.

7.4 *EASE Seminar Skills 1: Presentations*

A description of the materials is given in Chapter One section 1.2.3. A structural chart with pictures of screenshots of the *EASE Volume Two Seminar Skills* is given in Appendix 1.1. A description of the units selected for closer scrutiny is given in Appendix 7.2.

7.4.1 *Flowchart of Seminar Skills Unit 1*

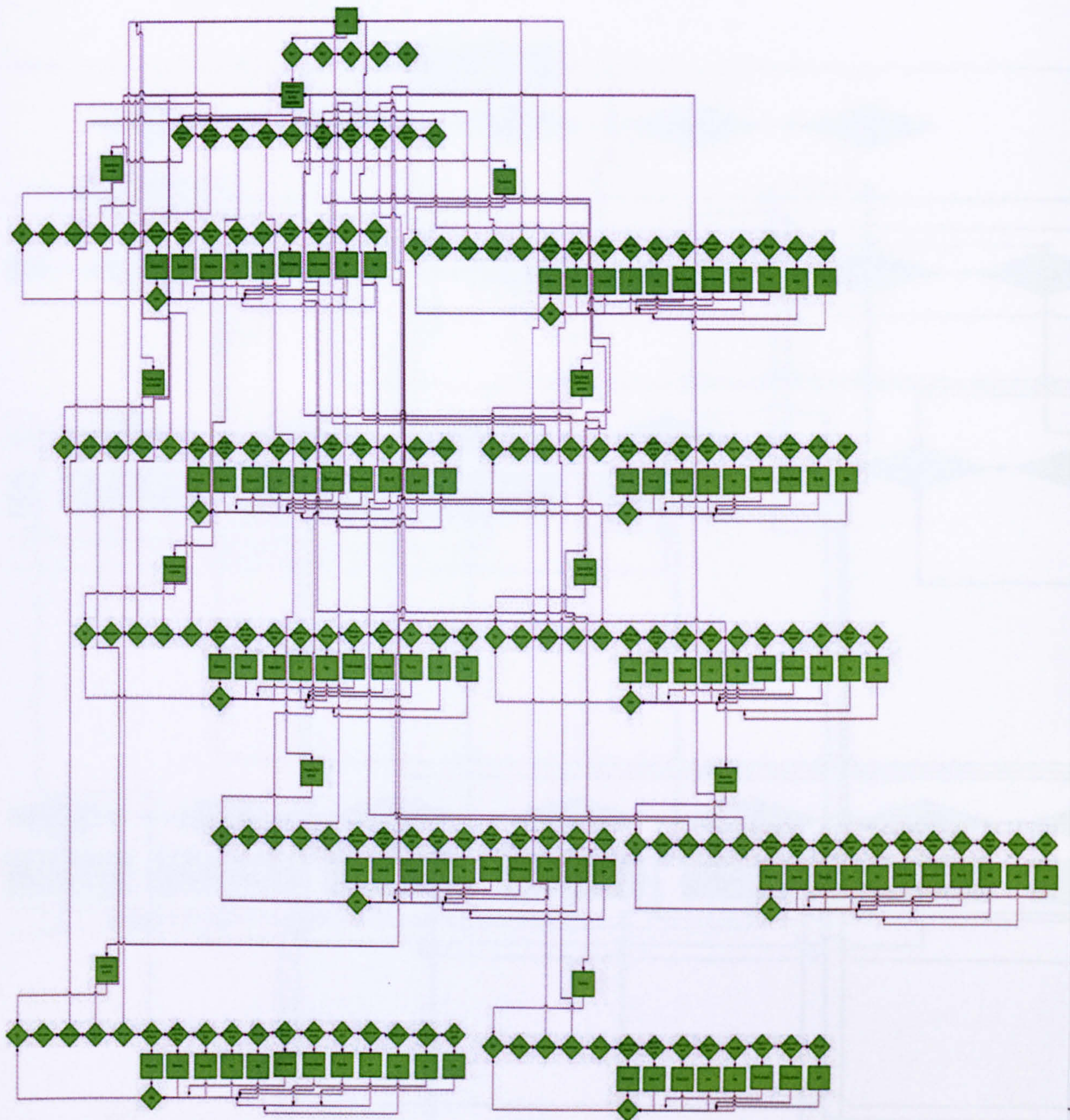


Figure 7.12 EASE 2 Unit 1 Flowcharts

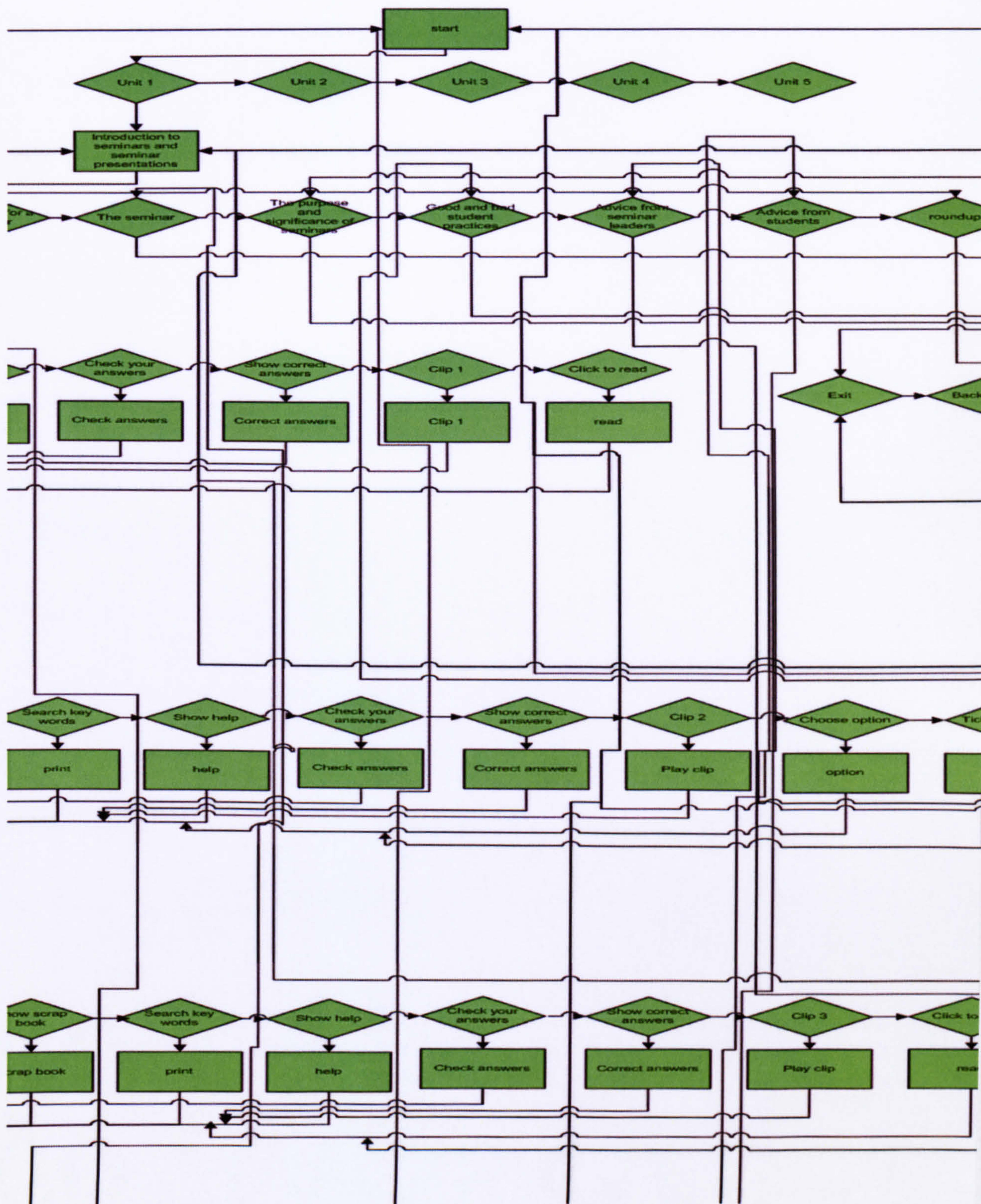


Figure 7.13 *Seminar Skills* Unit1: Closer Picture

7.4.2 Flowchart of Seminar Skills Unit5

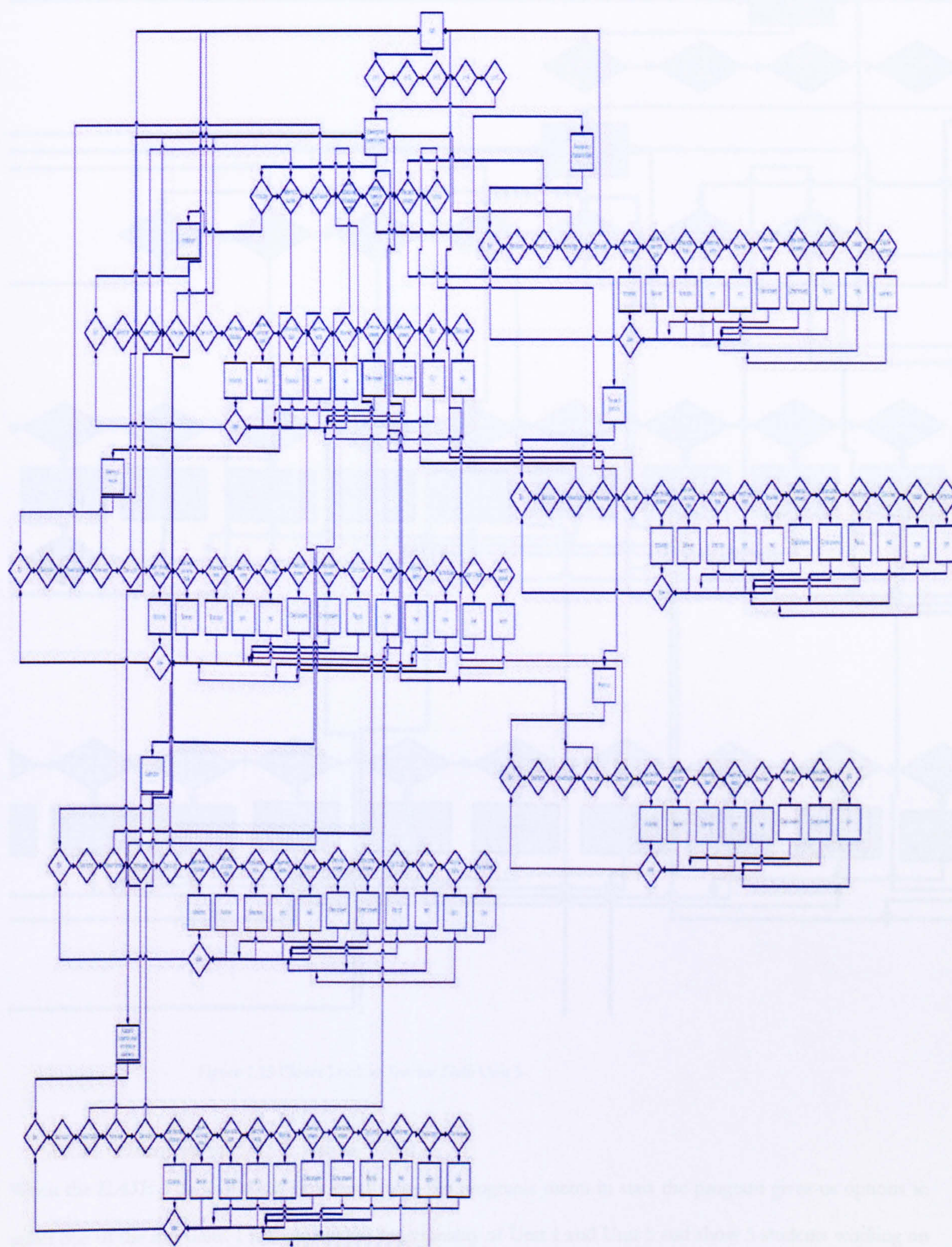


Figure 7.14 *Seminar Skills* Unit 5 Flowchart

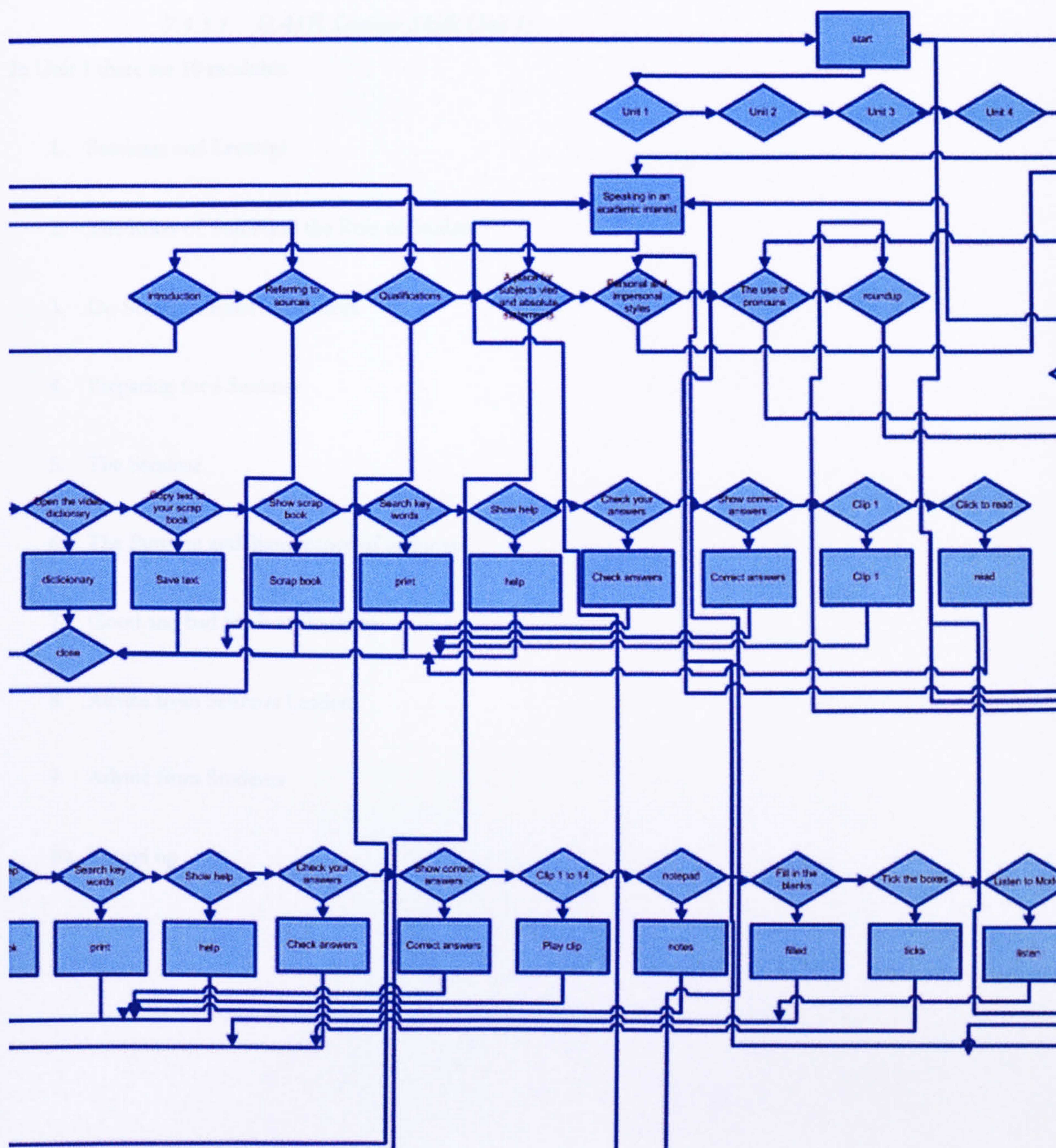


Figure 7.15 Closer Look at *Seminar Skills* Unit 5

7.4.3 Description of EASE Seminar Skills

When the *EASE 2 Seminar Skills* is opened from the programs menu in start the program gives us options to select one of the five units. I will explain the functionality of Unit 1 and Unit 5 and show 5 students working on each. Complete descriptions of units are given in Appendix 7.2

7.4.3.1 *EASE Seminar Skills Unit 1:*

In Unit 1 there are 10 modules:

1. Seminars and Lectures
2. The Roles of Tutor and the Role of Student
3. Do Seminars Exist in Sciences
4. Preparing for a Seminar
5. The Seminar
6. The Purpose and Significance of Seminars
7. Good and bad Student Practices
8. Advice from Seminar Leaders
9. Advice from Students
10. Round up

7.4.3.2 Flowchart of individual student working on Seminar Skills Unit 1

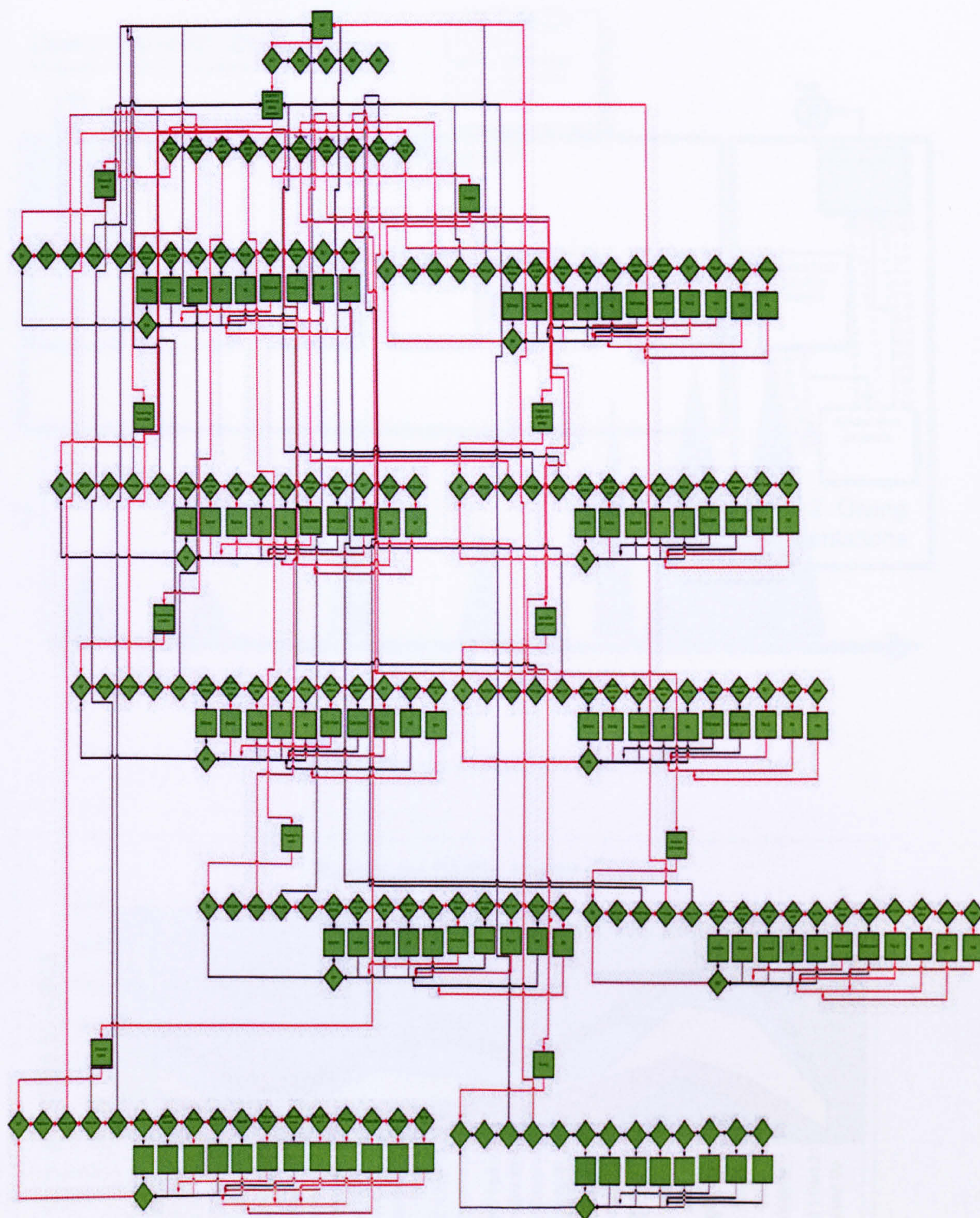


Figure 7.16 PC30 working on Unit 1 Seminar Skills

7.4.4 Five Students navigation through Unit 1 of Seminar Skills

Unit1: Seminar Skills

Navigation Patterns of Students

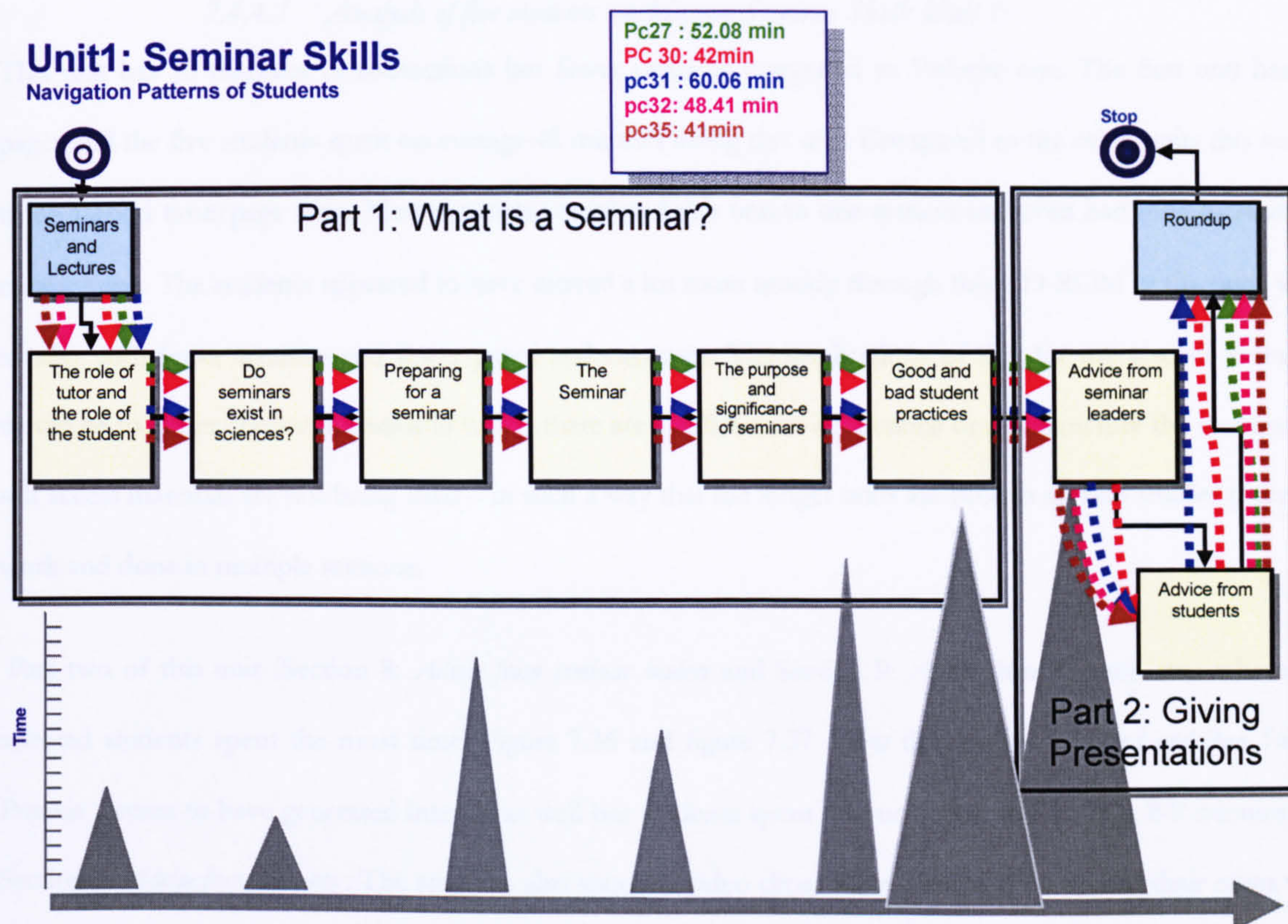


Figure 7.17 Navigation picture Unit 1 *Seminar Skills*

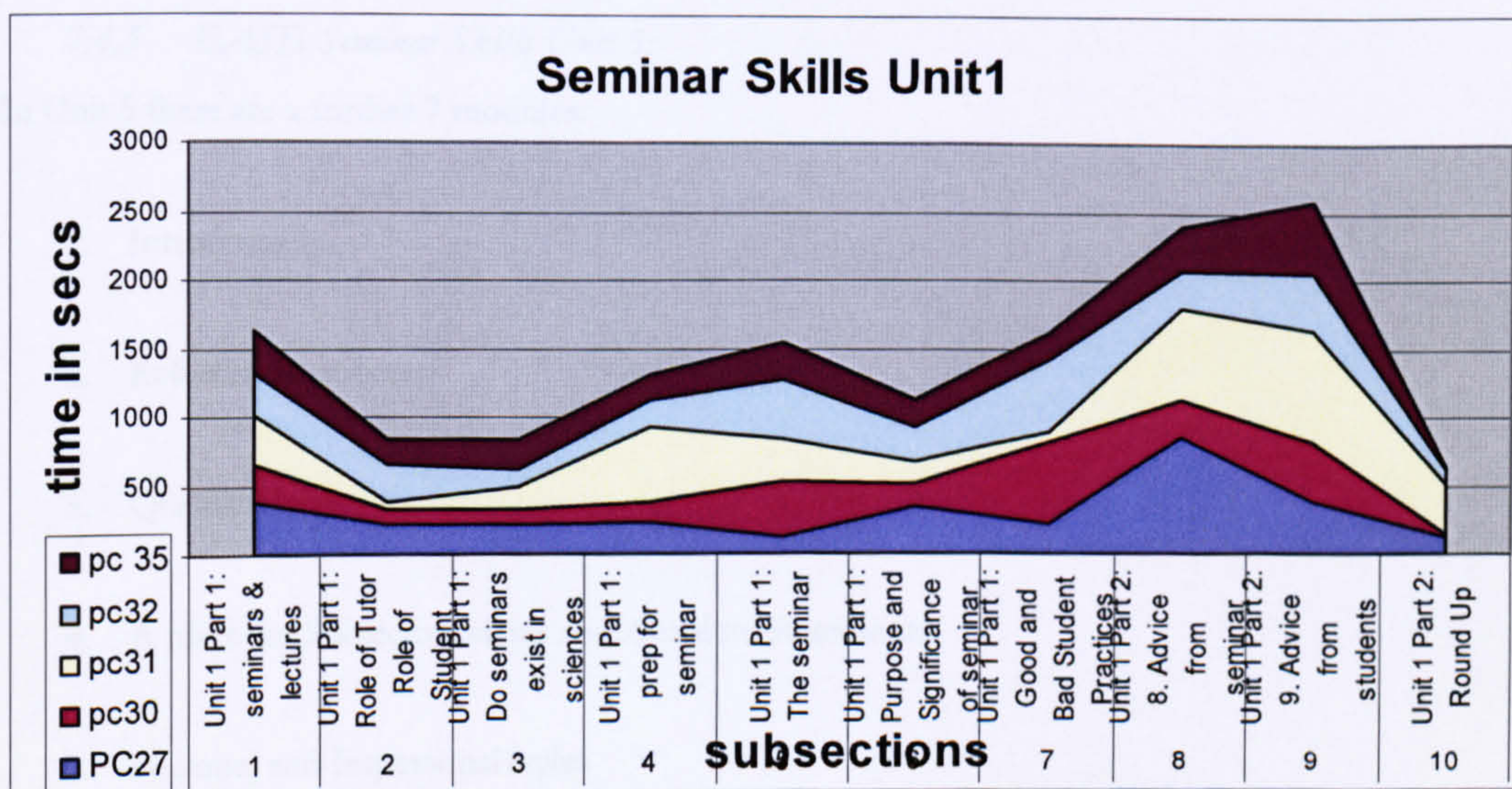


Figure 7.18 Time Spread graph of Unit 1 *Seminar Skills*

7.4.4.1 *Analysis of five students working on Seminar Skills Unit 1:*

This unit has 10 modules or subsections but fewer pages as compared to Volume one. The first unit has 27 pages and the five students spent on average 48 minutes doing this unit. Compared to the other units this seems to be a good time/page ratio. Most students completed this unit in one session and even had time to start the next session. The students appeared to have moved a lot more quickly through this CD-ROM as the units were smaller with fewer sections and fewer pages and exercises. The implications of this for the teacher evaluator would be to either choose materials in which there are smaller units or sections or else structure the sessions—if self access materials are not being used – in such a way that the longer units are broken up into smaller pieces of work and done in multiple sessions.

Part two of this unit (Section 8: *Advice from seminar leaders* and Section 9: *Advice from students*) was where the selected students spent the most time. Figure 7.36 and figure 7.37 show this. *Section: 7 Good and Bad Student Practices* seems to have generated interest as well but students spent the most time (on average 8-9 minutes) on Section 9: *Advice from students*. The students also watched video clips, made notes and compared their notes with those of the authors. Other fill in the blank exercises, true/false, and classifying information exercises were also attempted quite diligently. The students appeared to get more correct answers, which may help to explain why they preferred this volume of EASE.

7.4.5 *EASE Seminar Skills Unit 5:*

In Unit 5 there are a further 7 modules:

1. Introduction
2. Referring to sources
3. Qualifications
4. A place for Subjective views and Absolute Statements
5. Personal and Impersonal styles
6. The Use of Pronouns

7. Roundup

7.4.5.1 Flowchart of individual student working on Seminar Skills

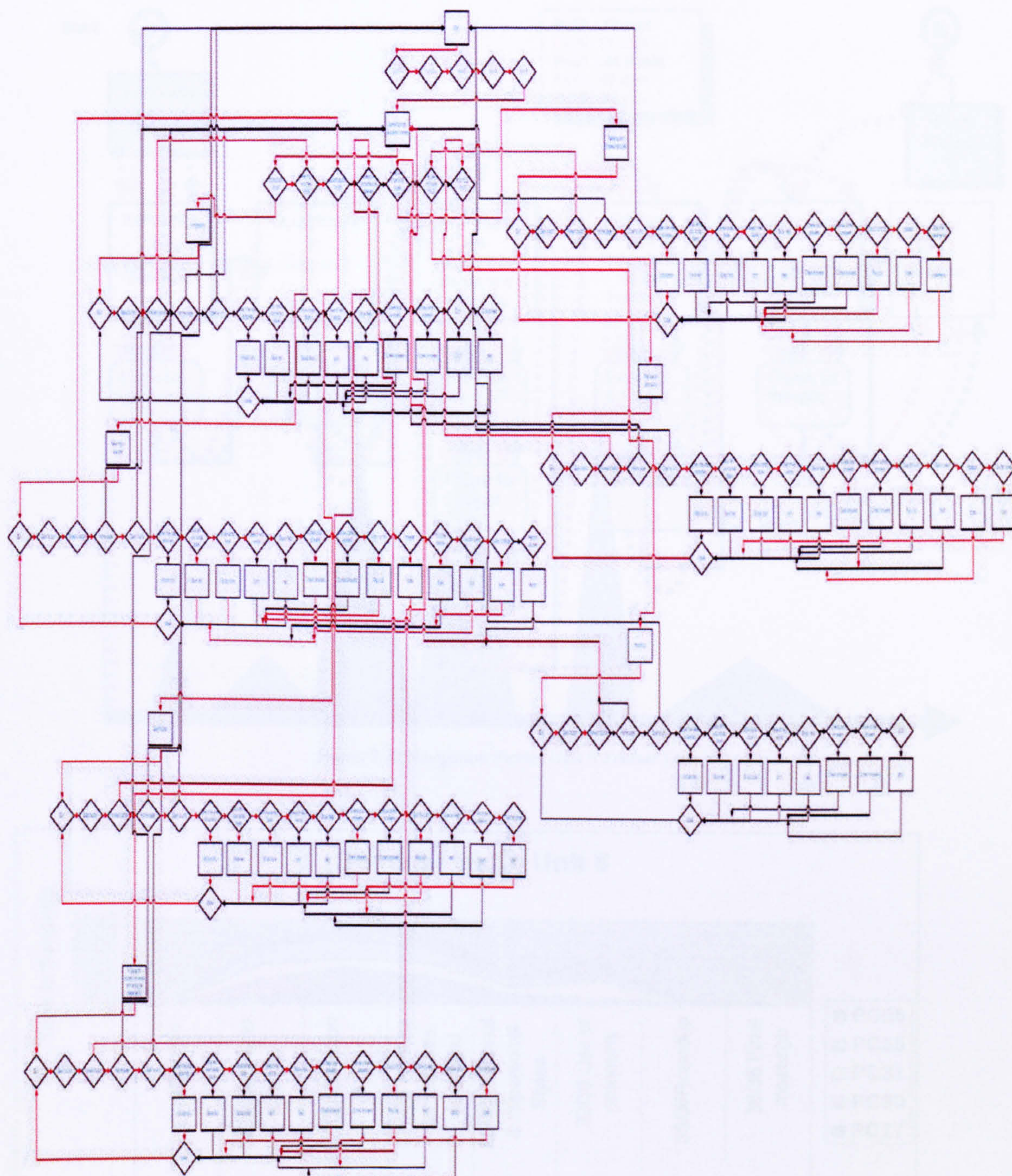


Figure 7.19 Student PC31 working on Unit 5 Seminar Skills

7.4.6 Five students' navigation through Unit 5 of Seminar Skills

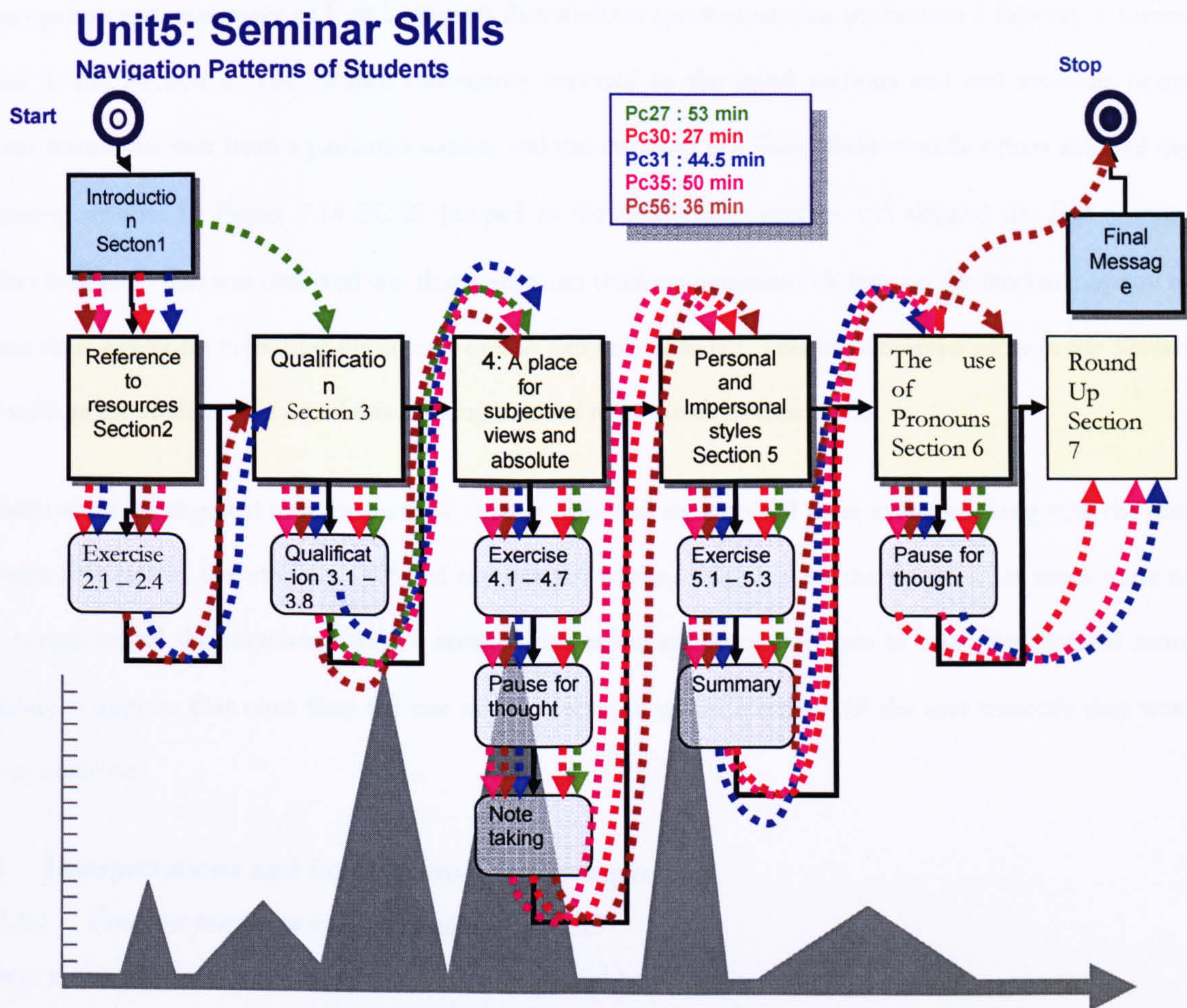


Figure 7.20 Navigation Picture Unit 5 Seminar Skills

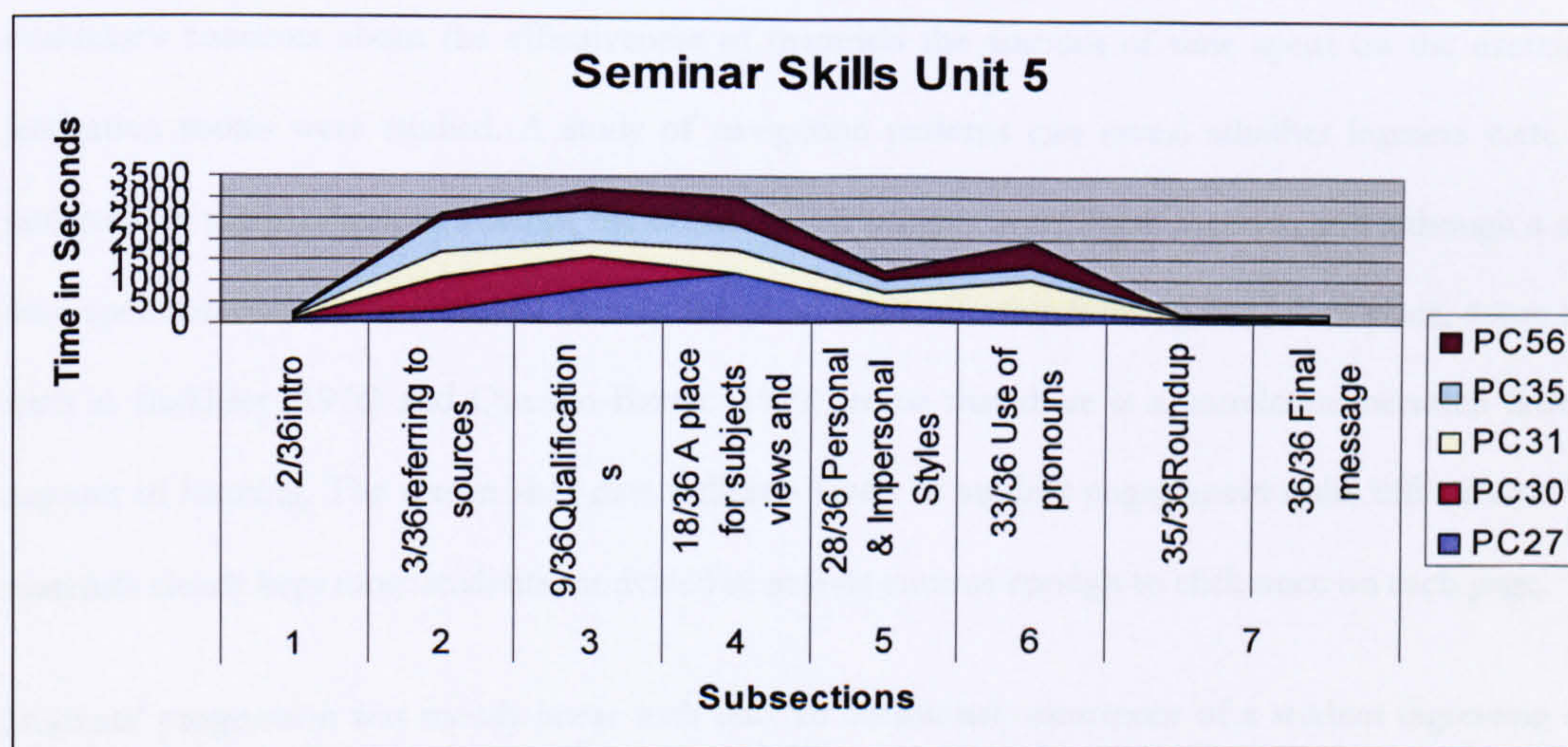


Figure 7.21 Time Usage Seminar Skills Unit 5

7.4.6.1 *Analysis of five students working on Seminar Skills*

The navigation and time usage of Unit 5 suggests that students spent more time on Section 2 *Referring to Sources*, Section 3, and Section 4. The greatest differences occurred in the initial sections and end sessions. Some students started the unit from a particular section and then worked in a linear fashion while others skipped the last closing session. In Figure 7.14 PC 27 jumped to the *Qualifications* section and skipped the first session. Another behaviour that was observed was that sometimes students continued clicking on the feedback option to increase their awareness even after the correct option had been selected. This was observed more in the *Seminar Skills* units as the students perceived this learning material to be closer to their needs.

Grammar work is integrated into the units of *Seminar Skills* and students did these exercises along with the rest. Although four out of five students (80% of the sample) worked diligently on the exercises, at times none of them completed all the exercises within a section (the earlier appearing exercises in a section received more attention.) It appears that once they did one or two exercises within a section of the unit correctly they were eager to move on.

7.5 Interpretations and findings from two perspectives

7.5.1 *From the perspective of the evaluation of EASE*

Answers to the questions raised in section 7.2 are attempted here.

Level of Engagement In order to determine the level of engagement with the materials and to answer the evaluator's concerns about the effectiveness of materials the amount of time spent on the exercises and the navigation routes were studied. A study of navigation patterns can reveal whether learners were sufficiently interested in the materials to attempt the exercises in a progressively linear manner, and although a study of the time spent on each section cannot directly reveal whether effective learning has taken place, some researchers, such as Baddeley (1976) and Quentin-Baxter (1999), argue that there is a correlation between time spent and amount of learning. The screen shot data indicates levels of student engagement quite effectively. The EASE materials clearly kept most students motivated or at least curious enough to click once on each page.

Students' progression was mostly linear with only an occasional occurrence of a student digressing or jumping directly to a section, but a pattern of not completing the unit right to the end was discernible suggesting a loss of interest or motivation. The flowcharts of EASE Volume 1 Unit 2 (section 7.3.5) seem to suggest this loss of

interest (although the sample may be inadequate to arrive at a definitive conclusion) this assertion about learner behaviour (particularly about Unit 2) can be made. One explanation for not completing the units could be that since Unit 2 was a long one they also may have left it to complete it another time.

Time Usage and Speed of Navigation In the earlier units of *Listening to Lectures* more diligence was observed and as the students became accustomed to the format and the novelty value of the experience diminished, they started moving faster. In unit 6 of Volume 1 the students seem to have spent most time on the *Antithesis*, *Synthesis* and *Notetaking* sections. Student PC 19 spent 25 minutes on the *Synthesis* section of Unit 6 and 35 minutes on the *Notetaking* section. The sections in *Listening to Lectures* were longer and therefore took more time than those in *Seminar Skills*.

The criterion of diligence can be upheld by the amount of time spent and by observing whether whole clips are being listened to or they are being skipped. Are exercises being done by and large, and what feedback is being consulted? An overall impression is that about 75-80% of the sample worked conscientiously and attempted 85% of the exercises in the units. When the students are speeding through the pages (spending 10-15 seconds per page) or not taking time to listen to the whole clip, it is difficult to make any claims that learning has taken place. "Although it often is not possible to determine the level of student engagement with the accessed material, information which is *not accessed* has *no opportunity* to be learned." (Quentin-Baxter 1999)

The calculation of time usage revealed that certain units were more demanding than others and students spent more time on these. Another aspect worth investigating was whether time usage revealed any correlation between design features and the students' level of engagement and motivated behaviour where they are completing exercises: The general impression about the second volume *Seminar Skills* was that the students found it lighter going and appeared to have got more answers right and also moved faster through the units.

Using the additional features such as the Scrap Book, Dictionary and Grammar book The additional features of the materials such as the Dictionary and grammar book were consulted by the students (see Figure 7.7 earlier). The dictionary was consulted more than the grammar book.

An evidence of learning effectiveness could be whether complete notes were written in the scrapbook and all the exercises were being done. The scrapbook was used by almost all students to make notes. The quality of notes was not assessed but notes were made.

Distraction Levels and Performance Tracking and monitoring data have been used to identify and study “outlier” behaviour in previous studies (Quentin-Baxter 1999). Within a particular study outlying behaviour can be identified and labelled but no generalizations can be drawn as there are so many variations within the normal ranges of behaviour. One example of outlying behaviour would be engaging in other activities at the same time as the specified task. There is some evidence in the data of students accessing multiple computer applications. For example, a student working on PC13 (not in the selected Sample) regularly checked the cricket score, but only for 10 seconds or so at a time, as the time bar on the top left corner indicates in the following two screen shots. He appears to check the scores as a reward for an activity done well, an interesting strategy that may have implications for the management of online learning generally

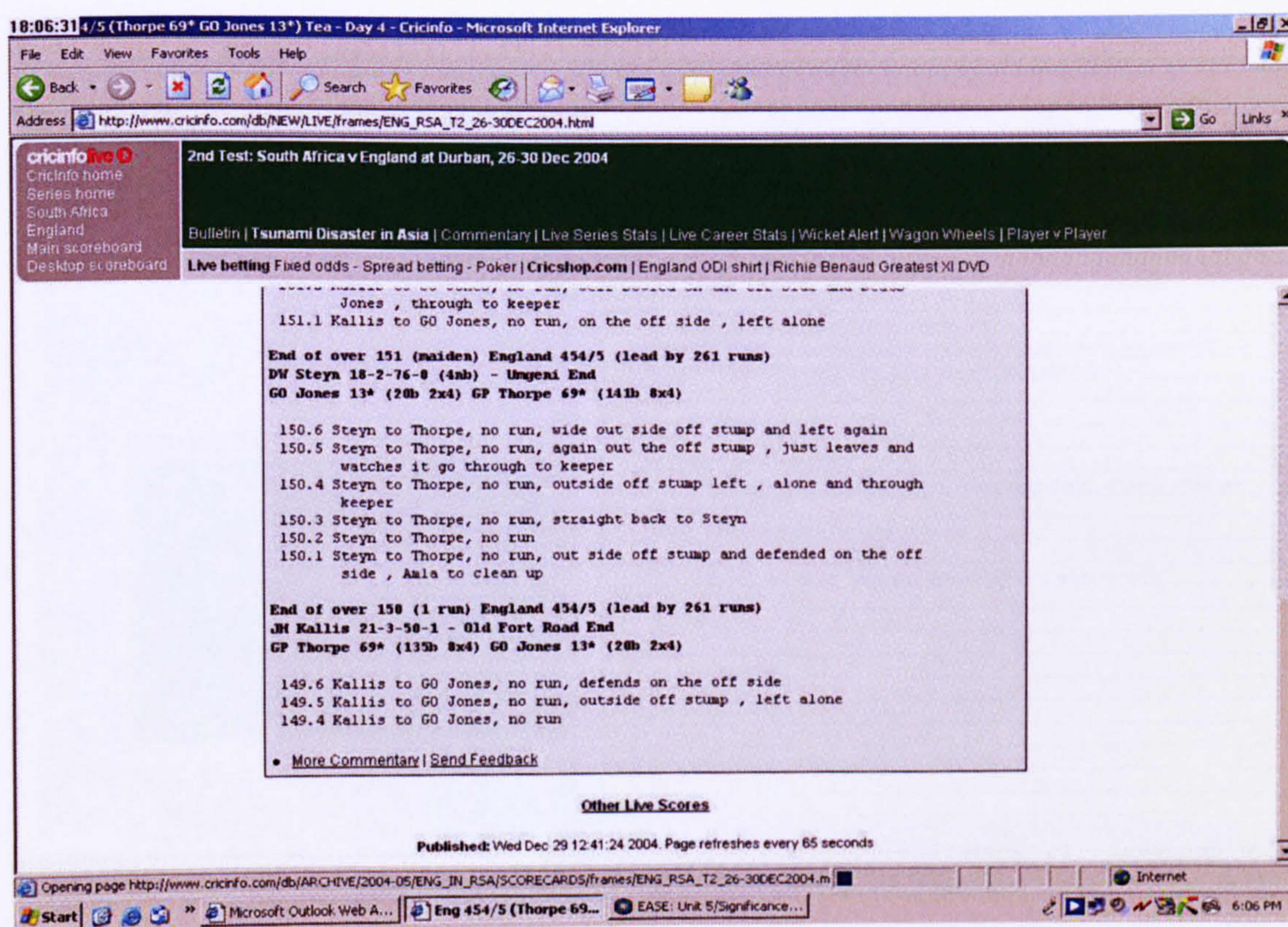


Figure 7.22PC13 Checking cricket scores. Check clock 18.06.31

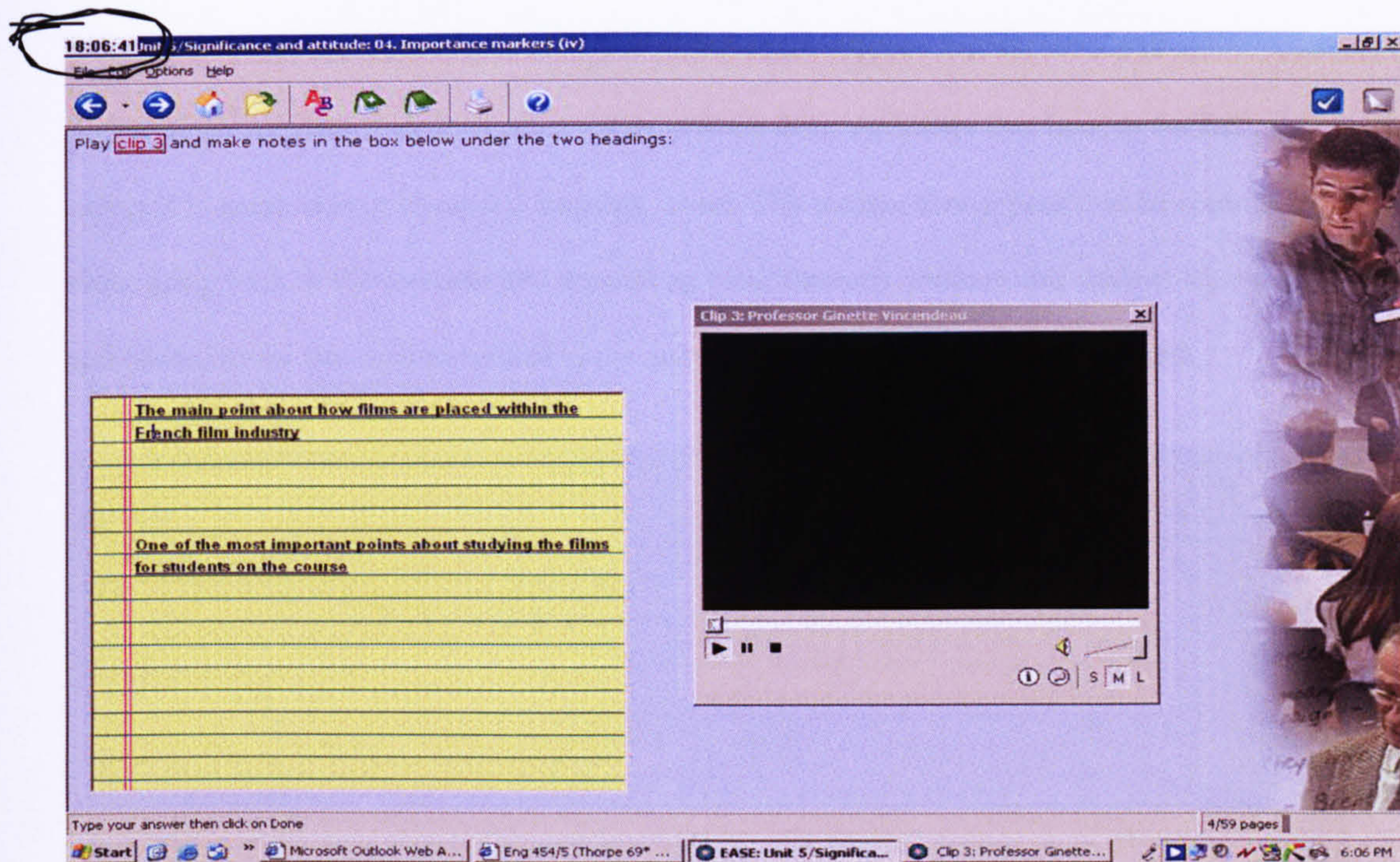


Figure 7.23 Pc13 Rewarding with cricket scores. Check clock 18.06.41 (10 sec interval)

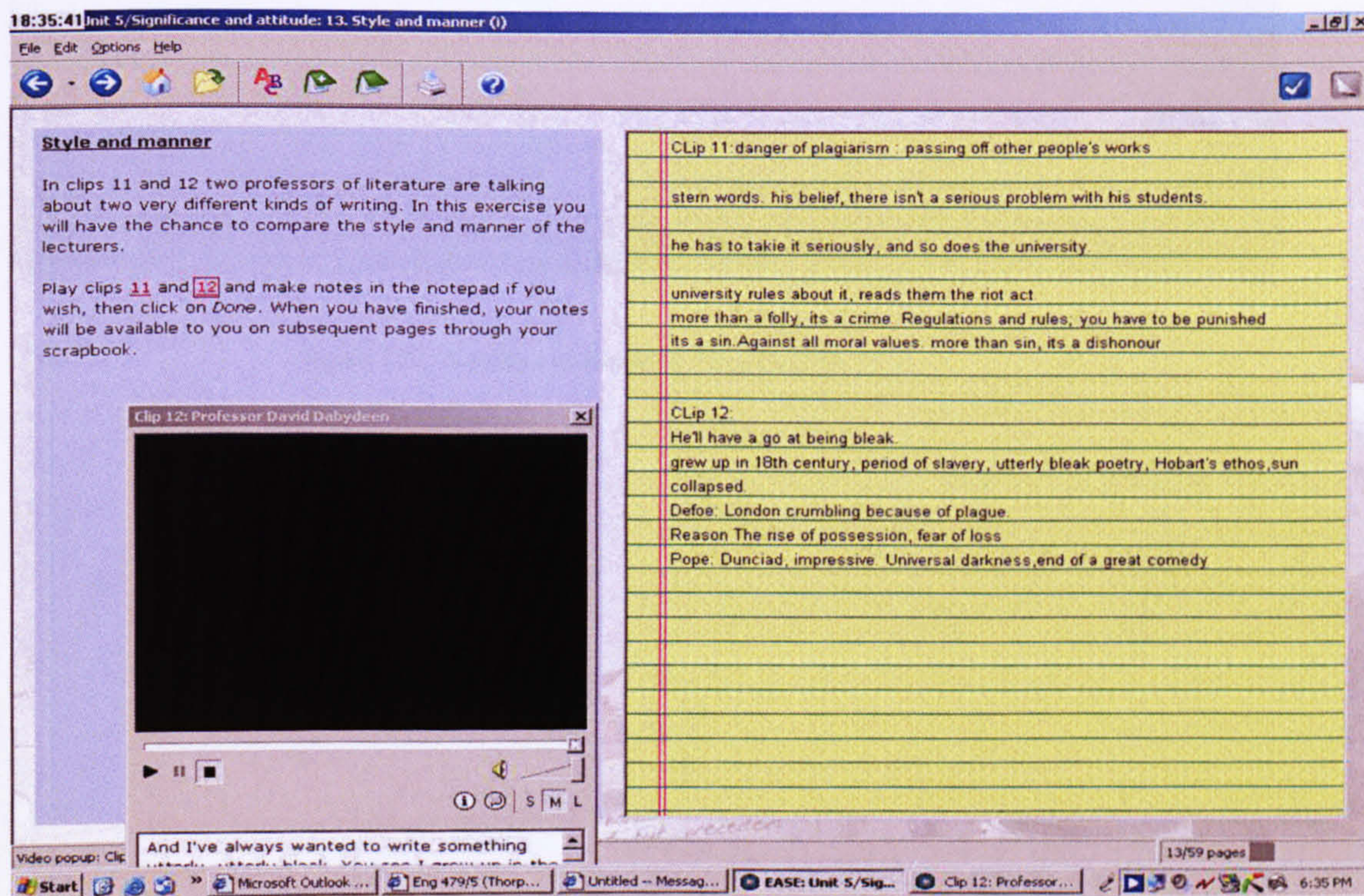


Figure 7.24 Screenshot showing PC13's note-taking effort

It can be seen that this student took complete notes, which suggests that his behaviour did not interfere with his performance. Another student accessed online journals from the library data base, as the following screen shots, captured in succession at 10 second intervals, reveal. This student also appeared to be committed to the EASE tasks, going back to the exercises and attempting them earnestly. Perhaps this student was under time pressure and needed to use this time connected to the university LAN to do some library research.

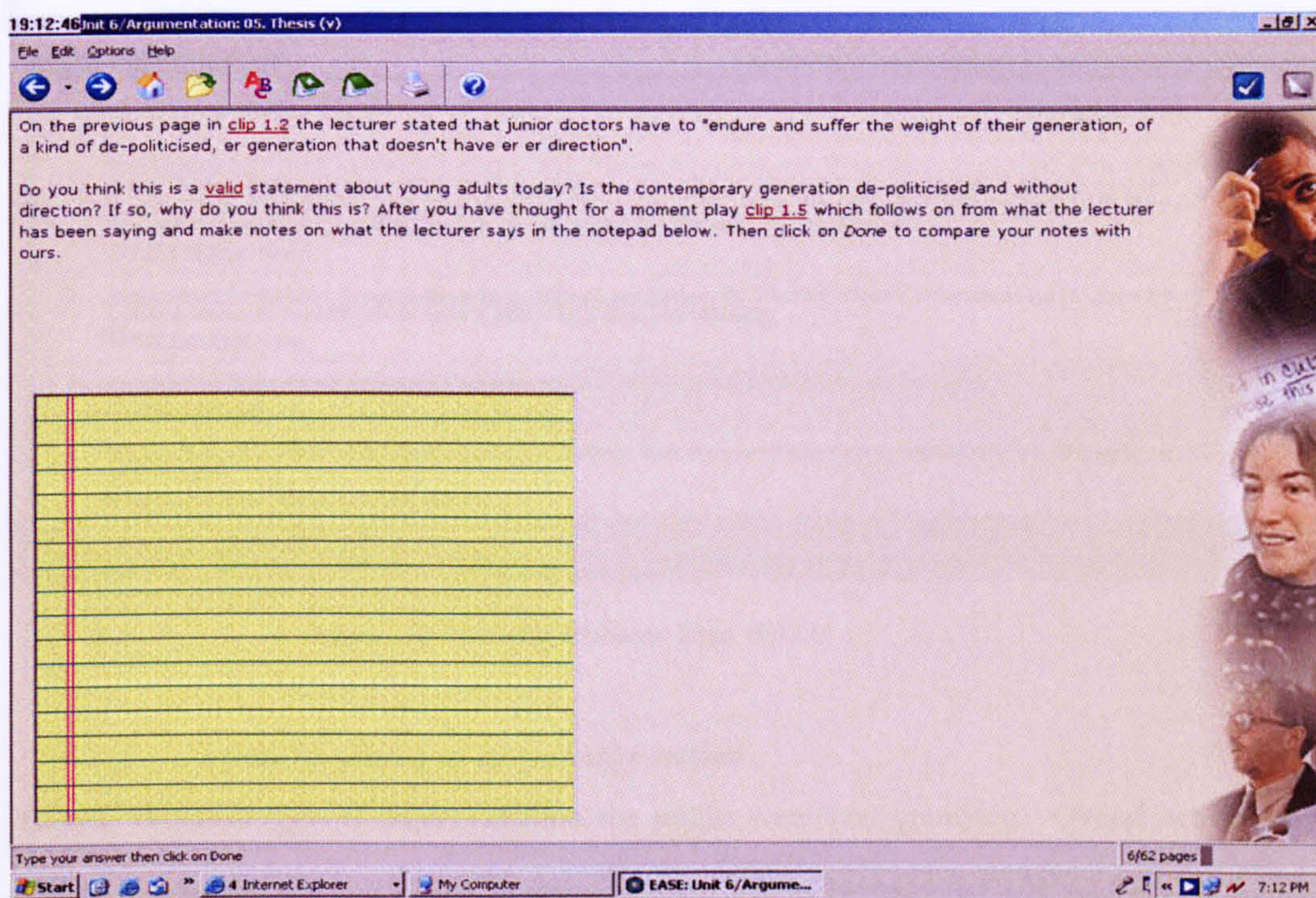


Figure 7.25 : A student multitasking. Time 19:12:46

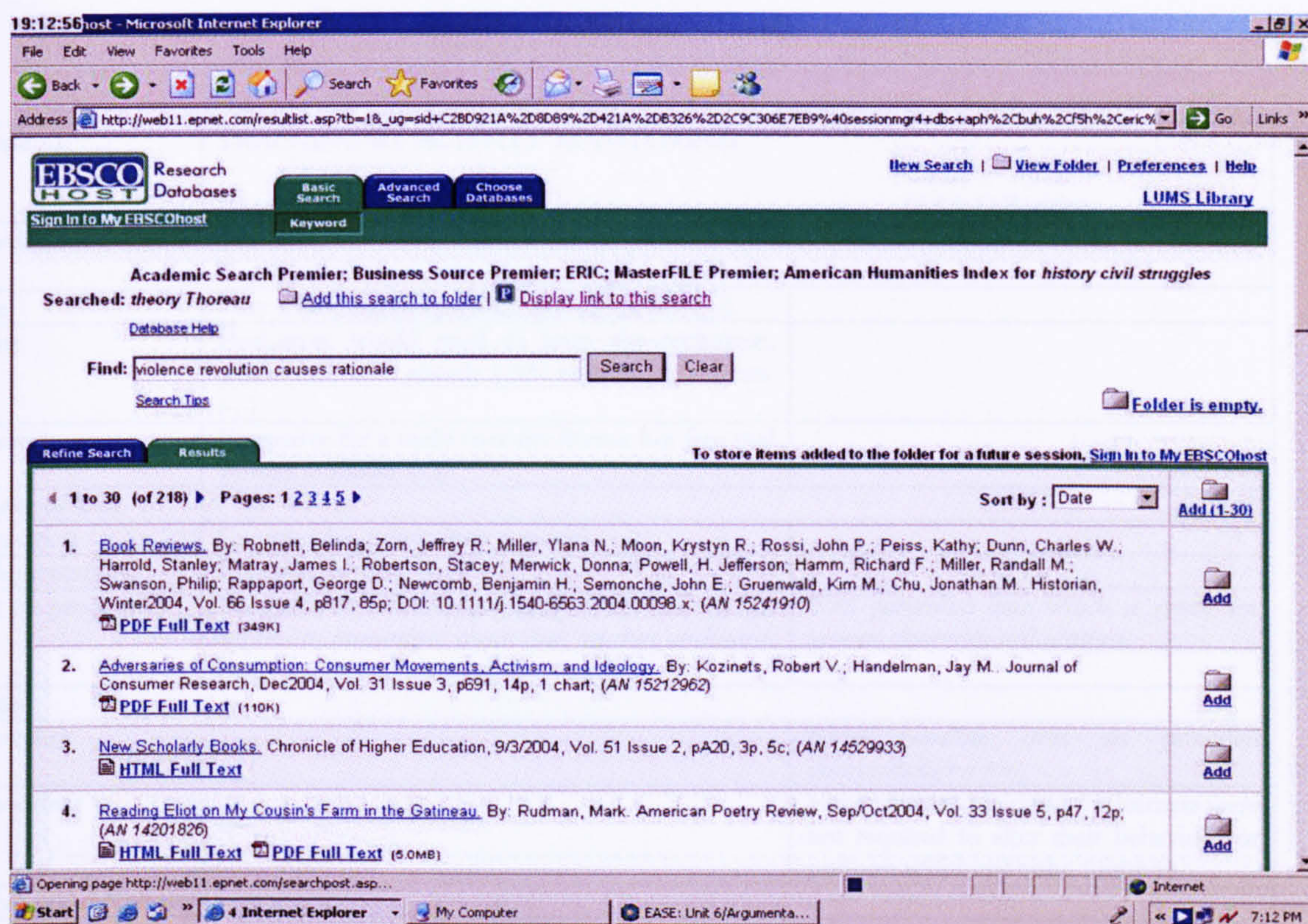


Figure 7.26 Accessing databases. Time: 19:12:56

7.5.2 Activity Monitoring as an evaluation method

Despite the challenges of implementation the results were very gratifying. Overall activity monitoring is a difficult method to use if a report of the evaluation has to be prepared and evidence of findings has to be presented to validate recommendations. But for the purposes of a teacher evaluator who does not have to persuade anyone and just needs information to act on it is a rewarding observation method which provides solid evidence of learners' interaction with the materials.

7.6 The method's performance on the criteria of evaluation

An explanation of the method's performance on the core criteria is given in Table 7.2. The two columns suggest negative and positive aspects of the method. The rating scale gives the degree to which the method is negative or positive (presented through +/– symbols).

Rating Scale: Degree expressed by the number of positive or negative symbols
Negative: (-,-,-) very poor/ (-,-) poor/ (-) slightly poor
Positive: (+) ok/ (++) fair /+++good/++++very good

Table 7.2 Questionnaire's performance on Criteria

Criteria	TRACKING BY ACTIVITY MONITORING	
	Negative (-,-,-)	(+++++) Positive
1. Cost Effectiveness		
Time	Very time consuming to set up and operate (-,-)	
Effort	Evaluators would need to learn the technique, unless they were already LAN networking experts. (-,-)	
Money	Expensive for a multi user site license but free trial version available. (-,-)	
2. Ease of Use		
Preparation	Considerable preparation required. (-,-)	
Implementation	Very problematic. (-,-)	
Data preparation	Data reduction for analysis purposes is difficult but applies to researcher more than teacher evaluator. (-,-)	Well presented data which is ready for general observational analysis.
3. Bias		
Researcher		None possible over an extended observation(++++)
Respondent		None observed because the learners were not required to alter their behaviour or perform additional tasks. (++++)
4. Ecological Validity		
Researcher Intrusiveness		Could be possible but the possibility of it happening over a period of time is unlikely(+++)
Environment Intrusiveness	Conditions of use and motivation levels may influence behaviours, but the method would capture it. (-)	The method captures it effectively (++)
5. Nature of Information		
Depth		Shows action rather than depth of thinking. (+++)
Objectivity		Screenshots provide very detailed information. Learner behaviour most objectively recorded without passing through any lens(++++)
6. Immediacy of Response		
Immediate/delayed		General impression can be gathered at first viewing of data without requiring extensive preparation. (+)ok
7. Usability measure		
Satisfaction	Shows the learner's involvement but does not overtly measure satisfaction. (-,-)	
Effectiveness		Learning effectiveness and usability are made evident by the method. (++++)
Efficiency		Problems with software can be picked quite well (+++)
8. Robustness of Method		
Validity		Very High because it measures what it is supposed to measure (++++)
Reliability		Factual information will replicate well. (++++)
Productivity		Highly productive vast amounts of data (++++)

Tracking can be done in any number of ways. Especially designed software which can lift data for evaluation can be created, or WebCT can be used for this purpose (Hwu 2003). This study used spy software which, if a multi-user license is purchased, can be expensive. Learning to use this program requires time and effort so on the scales of *cost effectiveness* this method does not score well.

This method also scores poorly on the scale of *ease of use*. Considerable preparation is required to install the software. This has to be followed by regular daily/ hourly maintenance of the software to ensure its smooth functioning. All this requires considerable expertise. Unless the user is an expert or expert help is on hand, collecting tracking data for observing learner behaviour can be difficult. Once the data has been collected preparing the data for interpretation is an even harder task. The data on its own reveals significant information of great value to the evaluator. However, to reduce the data for purposes of discussion is an arduous and time consuming task requiring special skills. Ideally a program to analyse data from tracking software should be used, but for this a programmer would have to be engaged.

There is no scope for any *bias* either on the part of the researcher or the respondent, therefore activity monitoring scores very positively on the scale of *bias*. If the respondent is biased about certain aspects of the materials, this will be reflected in patterns of use. There is no possibility of researcher bias unless in the final interpretative stage of analysis. The respondents' awareness that this data is going to be analysed eventually by the researcher might affect their working on the materials, but this awareness would be difficult to sustain over the whole duration of the course.

On the scale of *ecological validity* this method fares well as it is not an intrusive method. The students are likely to become oblivious of the fact that their work is being recorded as the activity monitoring works stealthily in the background. There is no researcher intrusiveness. However respondents' moods and motivation and the environment of use (for example if it is too noisy or cold or respondents are tired because it is a late session) can affect the data. These will be influencing factors across the board with all methods. This method objectively records any intrusiveness and therefore it can be accounted for in gauging the validity of the data.

On the criterion of *nature of information* this method reveals both deep and surface information. It is deep because it is very detailed. Each and every move of the learner is recorded and can be interpreted in as much detail and

depth as the researcher desires. The information is surface in the sense that actions (which are open to deep interpretation) are being recorded and not the thought processes of the learners working on the materials. It is a very objective method because it depicts actuality.

On the scale of *immediacy of response* this method gets both a negative and positive score. A quick impression of the learners' interaction with the materials can be gathered if just one respondent's work is looked at. If twenty respondents' recorded interaction is replayed and observed it can be time consuming and the results of the evaluation can be delayed. An impression of how the learners worked through the materials, whether they worked conscientiously and learnt anything can be determined by the amount of time they spent on each activity.

On the criterion of *usability measure* activity monitoring data does not show *satisfaction* directly because it is an intrinsic attribute and the method is not meant to capture it. It does capture a learner's involvement with the materials. *Effectiveness* of the materials is picked up by the method as the recorded interaction shows the levels of engagement between the materials and the learner and the number of times something goes right. Compared to *satisfaction*, *effectiveness* is an extrinsic measure and this method picks up both learning effectiveness and usability of materials. This method also makes it possible to judge the *efficiency* of the materials because if there are problems in the functioning of the software they become evident right away and screen shots capture it. For instance an error in the numbering of pages in Unit 6 of *Listening to Lectures* is picked up on all respondents' tracking data.

The method scores most highly on the criterion of *robustness* as the *validity* of the method is high. It captures the information that it is supposed to and the validity of the data collected through this method is high because it is entirely objective, recorded mechanically without any researcher bias or intrusiveness. The method scores highly on the *reliability* scale as well because irrespective of even the context, it will give the same kind of data when it is replicated. On the scale of *productivity* this method scores highly as well and quite a few discussion points emerge related to learner's behaviour, of which the constraints of this study do not permit a fuller exploitation, but which could be the subject of a whole thesis!

7.7 Summary

This chapter has presented data gathered from activity monitoring in different ways. It has shown navigation patterns and time spreads of a sample of students' work on selected units. The data was analysed from two perspectives: firstly, the teacher evaluators' concerns about the students' interaction with the materials were kept in view, and secondly the method's effectiveness was measured according to the effectiveness criteria. The teacher evaluator may find this method valuable because of the direct, irrefutable evidence it provides of learner behaviour.

With this chapter I conclude the discussion of findings of the four data chapters. The findings in relation to each other and the answers to the research questions are presented in the next concluding chapter.

CHAPTER EIGHT: DISCUSSIONS OF FINDINGS AND CONCLUSION

8.1 Introduction

This concluding chapter aims to synthesise the discussion of the previous chapters in the light of the main research question and sub questions. It starts with an overview of the study and its main objectives by revisiting the research questions. A summary of the findings relating to the performance of each evaluation method according to the core criteria will be followed by a comparative discussion of the four methods used in the study. The implications of these findings for methods of evaluating MM CALL materials, and the contribution that this study makes to the field, are also discussed.

8.2 An overview of the study

The fast pace of developments in multimedia materials and learning technology creates a need for evaluation procedures and processes to develop rapidly as well. This general principle applies equally to the development and evaluation of CALL materials and the associated development of appropriate evaluation procedures and methods. Studies dealing with evaluation of MM language learning materials are few at present. There is a need for more evaluation studies to be undertaken if the field of learning through multimedia is to move forward (Chappelle, 1997; Levy, 1997, Yildiz and Atkins, 1993). This study has aimed to address these concerns.

The study investigated methods which could be effectively used in the evaluation of multimedia CALL applications. In order to operationalise these evaluation methods and observe them in action an actual evaluation study was designed. This involved piloting a selection of methods of evaluating MM applications and selecting four for detailed investigation. The methods selected were Focus Groups, Retrospective Protocols, PLUM and SUMI Questionnaires and Activity Monitoring (screen capturing and keystroke logging).

To determine the qualities and limitations of these evaluation methods, a set of criteria was developed with reference to the literature on software evaluation methods in Human Computer Interaction (HCI) and the broader literature concerning methods within evaluation in education in general, and in the evaluation of

language teaching materials in particular. The data collected was analysed from two perspectives: firstly, from the point of view of the evaluator of the application/ materials, who is interested in the merits of the materials and their suitability for student use; secondly, from the point of view of this research, which explores the 'goodness of fit' between methods of evaluation in general and the specific requirements of summative, pre-purchase evaluations of CALL MM applications. The research design that emerged theoretically positions the work as a case study within which in-use and post-use evaluation data were gathered empirically from course enrolled students. Data collection and analysis used both qualitative and quantitative methods. Participants in this research were 12 ESL students from the University of Warwick and 40+ Freshman/sophomore students from a university in Pakistan. They were engaged in using the EASE CD-ROMs *Listening to Lectures* and *Seminar Skills-1: Presentations*.

Since the aim of this study was not to evaluate the materials but the evaluation methods, the findings are primarily an assessment of how the chosen methods performed according to the set of criteria developed as part of the research method. In the next section the findings are discussed, keeping in view the research questions of the study.

8.3 Revisiting the research questions

Drawing on the work reported in the previous four chapters, I shall revisit the sub questions first, in order to develop an answer to the main one. Each sub question relates to one evaluation method, and I shall summarise the advantages and disadvantages of each in relation to the criteria.

8.3.1 Research question 1.1

What are the qualities and limitations of Focus Group Interviews for evaluating multimedia CALL software?

Two focus group interviews were conducted in this study. This was the only method used which led to collective findings, whereas the others all drew solely on individual responses. The findings confirm that focus groups are a quick and efficient method of gathering evaluative opinion when compared to other methods of the study. They provide immediate results without the need to resort to extensive data preparation, and enable consensual negotiated positions to emerge. (A complete discussion of the findings based on the core criteria is given in Chapter Four Section 4.3.4).

Opinions were refined through debate and the participants tried to keep in view the presence of others in the group, and work towards consensus. This method did not provide much specific information about the usability of the materials, however, and responses were somewhat subjective, even though the participants consciously tried to retain some objectivity. Focus group moderators have to be properly trained and the teacher researcher's presence could influence the validity of the data (by creating a 'Hawthorne Effect' a phenomenon like 'reactivity' which is discussed in Cohen, Manion and Morrison 2000).

In sum, focus group interviews in my experience are a quick and easy method; their most important strength is that they provide a quick overall impression of the MM materials being evaluated. This is significant, since inability to gather an overall impression has been cited as a weakness of MM materials evaluations (Bayram and Nous 2004; Wilson 2000; Reiser and Kegelmann 1994). Litosseliti (2003:21) and other practitioners have claimed that it is difficult to generalise from focus group data, because of the small number of participants. Hémard's (2004) use of focus group discussion seems to uphold Litosseliti's (2003) views; like the findings from his user-walkthroughs, his findings from his focus groups were very detailed, but were too diverse to provide a clear over view of the materials he was evaluating. In my studies, however, retrospective protocols and focus groups yielded very different types of data.

Another aspect of focus groups which the general literature on focus groups mentions but CALL evaluative studies do not highlight is the sense of ownership of learning which comes with the opportunity to 'reflect' post-use on the materials. This was evident in my students' reflective commentary on the connections between language and culture (Student G, for example, commented that the materials were "sprinkled" with "all kinds of local flavour" and that "English English is quaint and funny and quirky"). The students referred to in Hémard (2004) did not seem to be able to see the connections between language and culture as clearly as this. Perhaps this difference between my data and Hémard's could be attributed to the different design and content of our materials.

The most important disadvantage of focus groups, perhaps insufficiently emphasised in the literature, is that the moderator has to be highly skilled in the art of moderation. He/she has to stay consciously objective and impartial, which may be difficult for teacher evaluators who are insiders to the context and may have strong opinions about what is beneficial for their students. (Cf: Chapter 4 Section 4.4.4).

8.3.2 Research question 1.2

What are the qualities and limitations of Retrospective Protocols for evaluating multimedia CALL software?

Two kinds of retrospective protocols; oral think-aloud protocols (ROP) and reflective emails (RRWA) were used in the study. These methods revealed varied, detailed and specific information which was more subjective and affective in nature than that derived from the other methods. A quick overall impression of the materials is possible through this method without having to resort to any kind of data preparation. However for purposes of research data preparation was required. The qualities of the method are that it is very easy to operate and is highly productive, gathering detailed opinion over an extended period of time. It is individualistic and learner centred. (A complete discussion of the findings based on the core criteria is given in Chapter Five Section 5.4.1).

My experience of using this method to provide a finely etched picture of the students' interaction with the materials seems similar to Hémard's (2004) experience of using user walkthroughs to explore 'mental models'. He found that his data from user walkthroughs was too narrow, however, because the students' comments only focussed on the task or activity they were engaged in at the time of verbal reporting. This phenomenon was also observed in my ROPs, but I was pleased with the highly focussed comments because I wanted to know what the students thought about the tasks, and I was not exploring mental models (like Hémard 2004) or learning strategies (like Legenhausen and Wolff 1990).

It must be borne in mind, however, that the ROPs were formatively collected while the learners were still in the process of working on the materials, and before they had had the opportunity to reflect on their impact. Learners' moods and motivational levels may differ on different occasions, and to a certain extent their responses may depend on external factors unconnected to the qualities of the materials themselves, limiting thus the reliability of any findings.

This method also did not capture the usability aspects of the applications as efficiently as had been hoped because the students were engrossed in the discussion of the efficacy of the pedagogical content of the materials rather than the functionality of the software. This however augurs well for an evaluator who is gathering data from the learners' perspective on the effectiveness of activities and tasks. If the objective of the evaluation is to find out how learners interact with certain tasks and activities, then this method will be highly effective.

8.3.3 Research question 1.3

What are the qualities and limitations of questionnaires for evaluating multimedia CALL software?

Two different kinds of questionnaires, *PLUM* and *SUMI*, were used in the study. Whereas all methods used were aimed at discovering the learning effectiveness and users' opinion of the materials, these were questionnaires directly focused on the usability and ergonomic dimensions of the application. Both retrieved specific information about the learner's experience, but one questionnaire provided quantifiable data while the other questionnaire served to corroborate this data by more qualitative means. The *SUMI* usability questionnaire used statistical analysis to report on the usability and related dimensions of the materials and gave quantitatively established findings. This method provided valid and reliable information on the materials. The information was quite easily derived as *SUMISCO* (a statistical analysis tool especially designed for *SUMI*) was used to interpret results. The findings confirmed that pre existing, well-constructed questionnaires like the *PLUM* and *SUMI* can be used by teacher evaluators to good effect. (A complete discussion of the findings based on the core criteria is given in Chapter Six section 6.5).

The most positive features of this method are convenience, efficiency and the ability to capture information regarding usability issues. The downside of this method, as with retrospective protocols, is that to a certain extent the learners' responses may depend on external factors unconnected to the qualities of the materials themselves. The context of use, including the timing of the evaluation, impacts on the validity and reliability of the findings.

In my use of both questionnaires I did not face the issues of low response rate or sections of questionnaires being left blank. There were only a very few negligible instances of one item being left out by one, or at worst two, students, perhaps because of the importance of proficiency in English and computer literacy in the Pakistani academic milieu and the seriousness with which my subjects approached their studies. Hémard and Cushion (2003a) and Hémard (2004) used questionnaires as evaluation instruments, but their response rate was low. Hémard reported that "only half the students who filled the questionnaire had used CALL in class, suggesting poor attendance or irregular use ..." (2004:516). Nevertheless, data from those students who did use CALL was used by Hémard to establish a new student profile and adjust materials design accordingly, and both these studies were able to convey an overall picture of students using CALL materials. I concur with Hémard

and Cushion (2003a) and Hémard (2004) that questionnaires do provide a good overall impression of what aspects of materials students like and what works for them.

8.3.4 *Research question 1.4*

What are the qualities and limitations of observation through Activity Monitoring and tracking data for evaluating multimedia CALL software?

Activity monitoring and computer aided tracking technologies were used to monitor students' learning and actual use of MM materials in this study. Data from activity monitoring revealed that the majority of the respondents navigated through the application with serious intent in a linear fashion, spending on average a minimum of thirty minutes on those sections of the materials that they claimed to have worked on (when responding to interview questions, and completing questionnaires and retrospective protocols). Thus this method may contribute to the validity and reliability of other methods if used in combination with them. (A complete discussion of the findings based on the core criteria is given in Chapter Seven Section 7.6).

This method is a very objective method and scores highly on the criteria of *Ecological Validity* and *Bias*. It is also a very robust method in terms of validity, reliability and productivity (Cf: Chapter 7 section 7.8). From an evaluator /researcher's perspective, data from activity monitoring identifies a great number of issues regarding the learning effectiveness of the materials. The limitations of this method are that it is extremely resource-intensive in both work hours and financial layout. The financial costs are high if a multi-user license has to be bought for the spy software and installed on many computers, but if an existing program like WebCt is used to collect tracking data then financial cost can be minimized. However it remains human resource intensive as an effective system has to be devised to interpret the data, and interpretation itself takes many hours. The method lends itself to quantitative analysis but preparing data for quantitative interpretation is a challenging task, the difficulties of which are well documented in positivistic audit trail data studies (Alexander and Hedberg 1994; William and Dodge 1993). In this study the approach was more qualitative. Learning effectiveness was determined by the amount of time spent on each section, and this was used as a data interpretation strategy. The difficulty of making the data comprehensible for an evaluation report may deter a teacher evaluator from using this method. However the potential for research using this method remains exciting.

Beatty and Nunan (2004), Hwu (2003), and Smidt and Hegelheimer (2004) all used either tracking technology software, or video based computer software, or both, to study learner behaviour. Some of this software had in-built video capture and replay facilities, and some simply displayed screenshots (still camera shots as opposed to video). Beatty and Nunan's (2004) study, using Lotus ScreenCam screen capturing software to study collaborative behaviour, did not fully exploit the method because it had to be studied along with the video recordings from a video recorder (not built into the software) to get meaningful data. Hwu (2003) was primarily interested in differentiating the amount of time users spent on each page, but did not gather enough data to draw any definitive conclusions. Smidt and Hegelheimer (2004) used Camtasia recorder screenshots, but only as a corollary to qualitatively analyse post task interviews.

Perhaps the most interesting evaluation research using this method is that of Desmarais *et al* (1998), who studied learners' navigation patterns, first by discerning and drawing them from verbal protocols and then by designing tracking software which observed how learners navigated through materials. Desmarais *et al* distinguished between linear and chaotic patterns and reported that intermediate students presented more linear patterns than non-linear ones. Earlier work by Recker (1994), however, suggests that novices use more linear patterns and experienced learners prefer a more top down non-linear approach. My learners were experienced computer users but new to CALL MM materials like EASE, and their patterns were all linear (excepting one case, where a student jumped straight to the last unit (section 7.3.5. figure 7.10 student pc 24), but only in order to resume at the point where h/she had left off in the previous session). My learners did leave some exercises incomplete, but their route was always forward, with no backtracking except to the instructions page.

Perhaps Desmarais *et al*'s (1998) notion of linear and chaotic navigation patterns is less applicable to modern MM materials; which may have become more interesting in recent years, compelling students to move in a more linear fashion so that they do not miss out on anything. It is also possible that chaotic navigation patterns are more typical of first time users, who later, if they have understood the instructions correctly, adopt more linear routes. Another reason for why students may have come to understand that 'linear' is the best way to learn could be what Bax (2003: 23) refers to as 'normalisation' (the general ease that comes through familiarisation). Studying navigation patterns certainly seems to be a good means of keeping track of developments and changes in the learning behaviour of learners.

8.4 Towards best practice in MM materials evaluation

To answer the main research question of this study the methods will be discussed in relation to the evaluation criteria. This will be followed by a discussion of possible effective combinations of methods:

What is best practice in a learner-centred evaluation of MM CALL materials and which methods or combination of methods are likely to be most effective?

Since the aim of this study was not to evaluate the materials but to study the methods of evaluation, the focus in the findings is on how the chosen methods performed according to the pre-set criteria. Each of the chosen methods was selected as representative of one type used in the fields of HCI and educational evaluation. The discussion in chapter two (section 2.5 and figures 2.1, 2.2 and 2.3) highlights this and gives more complete information about the broad categories from which the four methods were selected. The methods within the domain of HCI which this study focuses on are *user-testing* (figure 2.3) rather than *expert analysis*. Moreover, the selected methods also represent the groups mentioned in figures 2.1 and 2.2. Educational materials evaluation methods and research methods used in the social sciences were combined with methods used in HCI for usability evaluation, to account for the computer -based nature of MM materials.

CALL MM evaluation studies cannot afford to ignore new developments in the software industry and HCI aspects which impact on the design of CALL materials. Ivory and Hearst (2001) give an indication of the scale of Usability Evaluation (UE) methods available in HCI; they refer to windows, icons, pointer, and mouse (WIMP) user interfaces (UIs) and Web interfaces, claiming that “we surveyed 75 UE methods applied to WIMP interfaces, and 57 methods applied to Web UIs” (2001:474). It is important for CALL evaluation to stay aware of these changes and developments in UE which utilise functional and technical advances in the field.

HCI usability evaluation takes into account the learning effectiveness of materials from an ergonomics/HCI theoretical perspective. The CALL materials evaluator has to combine the evaluation of pedagogical content with HCI usability evaluation principles. Usability evaluation methodologies, methods and techniques are an advanced and concentrated field of study, since the need for good evaluation practice and methods is more real, immediate and core to HCI. From a purely pedagogical perspective, developments in usability evaluation which test learning effectiveness of materials also have to be watched by CALL evaluators and developers.

The focus of this thesis has not been to look at evaluation methods from the perspective of the designers and developers involved in formative evaluations, where obviously the developments in HCI will have more sway. However, even for summative evaluation like the one undertaken in this study, developments in HCI usability testing and inspection have to be noted and adapted. In my view “best practice” in CALL materials evaluation has to incorporate methods from HCI which reflect state of the art best practice in usability evaluation. One integral feature then of best practice is keeping pace with developments in HCI summative evaluation.

Another important aspect of best practice is the inclusion of the end user. It is important to remember that the end users of CALL products, our learners, are also evolving alongside the developments in computer and internet based technologies. Their expectations from CALL materials may have grown with their increased exposure to highly sophisticated gaming and learning software. (A respondent’s comments in the focus group revealed this expectation when he compared EASE with TOEFL preparation materials. Cf: Chapter four, section 4.3.2.3.1 and Appendix 4.4). Methods which best elicit learners’ concerns and values have to be part of best practice.

In the above discussion I have attempted to answer the first part of the main research question. In the following section, I will attempt to respond to the second part of the question by looking at the performance of the selected methods in relation to the evaluation criteria.

8.4.1 Comparative performance of methods on the evaluation criteria

Triangulation of methods and validation of data from multiple sources is considered good practice in research because it makes the researcher confident about the findings and helps to overcome the problem of ‘method-boundedness’ (Cohen *et al.* 2000:113). According to the same principle, a combination of methods may best serve an evaluation. The usability evaluation literature also suggests use of multiple evaluation techniques to overcome the problem of each method only covering a subset of the possible routes or actions end-users may take (Dix *et al.* 2004; Nielsen 1993). To arrive at a combination of methods which leads to best practice in evaluation, a comparative analysis of the methods is undertaken in this section.

Hémard (2004) describes different methods that can be used by designers for the evaluation of CALL MM materials, but presents no criteria for measuring their effectiveness. Preece *et al.* (1994), Dix *et al.* (2004) and

other seminal texts in HCI, on the other hand, emphasise key attributes of usability evaluation methods. Their criteria are based on the HCI literature and combine partial lists of attributes from various HCI sources. My own set of core criteria is a particular instrument I have devised in order to gauge the potential of each of the methods I have reviewed. The set is derived from similar criteria in the HCI usability evaluation literature, but concerns expressed in the literature on research methods in education and educational evaluation have also been kept in view. I believe that the way in which I used the methods and developed the criteria makes them potentially appropriate for use by non-HCI practitioners and teacher evaluators working in the field of ELT MM materials with no specialist knowledge of HCI usability evaluation. A detailed definition of each criterion is given in section 3.8 in Chapter Three.

The chosen methods of this study are now assessed on the core criteria developed in this study. Table 8.1 presents a summary of the performance of the methods on the evaluation criteria, using the star rating scale which is explained in Box 8.1. Rating the findings from each method on the criteria was a comparative and value laden exercise. Yes/no, or high/low scales had to be calibrated with plus/minus signs to show their exact depth, value and polarity. The star rating scale was then introduced in order to facilitate comparison between methods. The star rating scale is ubiquitous and easily comprehensible and conveys meaning to exactly the right gradient.

Rating Scale for Table 8.1	
5 stars	★★★★★ Very Efficient/Good
4 stars	★★★★☆ Efficient/ Good
3 stars	★★★☆☆ Average/Neutral
2 stars	★★☆☆☆ Not so Efficient
1star	★☆☆☆☆ Poor

Table 8.1: Comparative Performance of Methods

Criteria	Focus Groups	Retrospective Protocols (1)Oral (2)Written	Questionnaire PLUM SUMI	Activity Monitoring
1. Cost Effectiveness				
Time	★★★★★	★★★★☆	★★★★☆	★★★★☆
Effort	★★★★☆	★★★★★	★★★★☆	★★★★☆
Money	★★★★★	★★★★★	★★★★★	★★★★☆
2. Ease of Use				
Preparation	★★★☆☆	★★★☆☆	★★★★☆	★★★★☆
Implementation	★★★★★	★★★★☆	★★★★★	★★★★☆
Data preparation	★★★★☆	★★★☆☆	★★★★☆	★★★★☆
3. Bias				
Researcher	★★★☆☆	★★★★★	★★★★★	★★★★★
Respondent	★★★☆☆	★★★☆☆	★★★★☆	★★★★★
4. Ecological validity				
Researcher Intrusiveness	★★★☆☆	★★★★☆	★★★★☆	★★★★☆
Environment Intrusiveness	★★★★☆	★★★★☆	★★★★☆	★★★★☆
5. Nature of Information				
Depth	★★★★☆	★★★★☆	★★★★☆	★★★★☆
Objectivity	★★★★☆	★★★★☆	★★★★☆	★★★★★
6. Immediacy of Response				
Immediate-Delayed	★★★★★	★★★★☆	★★★★☆	★★★★☆
7. Usability measure				
Satisfaction	★★★★☆	★★★★☆	★★★★☆	★★★★☆
Effectiveness	★★★★☆	★★★★☆	★★★★☆	★★★★★
Efficiency	★★★★☆	★★★★☆	★★★★☆	★★★★★
8. Robustness of Method				
Validity	★★★★☆	★★★★☆	★★★★☆	★★★★★
Reliability	★★★★☆	★★★★★	★★★★☆	★★★★☆
Productivity	★★★★☆	★★★★★	★★★★☆	★★★★★
Final Score	62/95	67/95	77/95	65/95

The total number of stars that can be attained is 95. The method that scores the highest on the nineteen criteria is the retrospective reflective written account, followed by the SUMI questionnaire. Although focus groups and activity monitoring do not score as highly they also have their value. Focus groups in particular did not fare well in this particular evaluation, but with proper training the moderator might produce more objective results, with

less evidence of bias. Similarly activity monitoring would be easier to use and more cost-effective if an automated data retrieval and interpretation program was available to track the data. The following section discusses the methods with reference to the criteria.

Focus group interviews were the most *cost effective* method of all the four methods, followed by retrospective protocols. The more quantitative positivist methods were costlier financially and in terms of human effort and time. The costliest method was activity monitoring. Questionnaires by themselves were not very costly once the right selection of questionnaire had been made.

On the criterion of *ease of use*, retrospective protocols, particularly reflective written accounts through email, scored most highly. Questionnaires were the second easiest followed by focus groups. Focus groups were the quickest means of obtaining an overall impression of the materials. Activity monitoring was the most difficult to use as a lot of expertise was required to maintain the software on the terminals and to keep it running for the duration of the course. Data preparation for interpretation is laborious if findings have to be presented in a report to a commissioning authority or administration. But for the teacher who just needs information to act upon, activity monitoring is a method with great possibilities.

On the criterion of *Bias*, activity monitoring scores very highly as it is the most objective of methods and at no stage can any bias affect data - not even "analyst bias" at the interpretation stage because the data speaks for itself. Respondent bias is inherent to a certain degree in all the other methods because opinions are being sought. Bias is perhaps at its highest in ROP think-aloud because spontaneous or concurrent views are being recorded. Researcher bias is only evident in the case of focus groups where the moderator, as an 'insider to the context', may inadvertently influence the discussion.

On the criterion of *ecological validity* all the methods indicate a certain degree of *environment intrusiveness*. The most ecologically valid method is activity monitoring, closely followed by questionnaires, where there is no researcher intrusiveness although a small degree of environment intrusiveness is possible.

The *nature of information* from the retrospective protocols was most specifically deep, subjective and detailed. Activity monitoring provided the most objective and accurate information, which was at the same time very

detailed and deep. Questionnaires and focus groups also revealed usefully pertinent information, making these methods equally effective on this criterion.

An important criterion for the teacher-selector/evaluator is *immediacy of response*. The method that delivers results most immediately is the focus group. The other method that delivers on this criterion is retrospective protocols. The results from the other two methods are delayed.

The method that delivers most relevant data on *usability* is the questionnaire, particularly SUMI which is especially designed as a usability measure. Data on the sub scale of *satisfaction* and *effectiveness* is elicited very well by retrospective protocols and focus groups. Activity monitoring does not capture direct *satisfaction* data but the recorded functionality of the software can be used to determine *effectiveness*.

The most *robust* method is activity monitoring because of its high *validity*, *reliability* and *productivity*. However because of implementation difficulties this method may not be the most practical. Retrospective protocols score well on the scale of validity. SUMI's validity and reliability has been rigorously tested and the inventory has been very carefully designed with many iterations. The reliability of all the methods except activity monitoring is to some extent context dependent however. The overall most practically robust methods are retrospective protocols and SUMI type questionnaires.

The above discussion of performance on the criteria was undertaken to determine best practice and observe the comparative strengths and weaknesses of the chosen methods. The next section deals with how this knowledge could be operationalised by the teacher-selector-evaluator in selecting a combination of the methods for conducting an evaluation.

8.4.2 *Different combinations of the model*

The selection of methods to be used should depend on the objectives and the focus of an evaluation. A combination will have to take into account the following considerations: what an evaluation method uncovers; what evaluation objectives are achieved; and what evaluation questions are likely to be answered by the combination of methods.

Table 8.2 gives in summary form some of the different combinations of methods that are possible.

Table 8.2: Combinations of Methods

Evaluation Objectives Learner Focus	Focus groups(FG)	Retrospective Protocols (RP)	User Questionnaires (UQ)	Observation: Tracking data Activity Monitoring(TAM)	Combinations
Recording learner's reactions to MM	√	√	√		FG+RP+UQ
Recording learner's attitudes to MM	√		√	√	FG+UQ+TAM
Recording skills developed in the learner	√	√		√	FG+RP+TAM
Recording knowledge gained by a learner	√	√			FG+RP
Evaluation Objectives : Usability Focus					
Gauging the ability of the learners to perform a task in a given context		√		√	RP+TAM
Gauging how easily the learners can carry out the task		√	√	√	RP+UQ+TAM
Gauging how quickly the learners can carry out the task				√	TAM
Assessing the overall quality and acceptance of materials.	√	√	√		FG+RP+UQ
Detecting faults and problems with the materials		√	√	√	RP+UQ+TAM
Determining ease of use of the product for a novice learner		√		√	RP+TAM

Table 8.2 suggests some possible combinations to guide a teacher-selector-evaluator in their selection of methods. For example, focus groups (FGs) successfully record all aspects or evaluation objectives in the learner focus section, whereas RP and TAM are perhaps better for recording usability aspects. If the objective is to record the skills developed in the learner then FG+RP+TAM can be used. If usability is the focus and the

objective is to gauge how quickly learners carry out the task then the method TAM may yield most reliable results.

A combination of methods for the teacher evaluator who is interested in a quick and easy in-use or post- use evaluation would be the SUMI questionnaire and focus groups. Reflective accounts by email are also a quick method of accessing learners' opinions and could be used in conjunction with questionnaires and focus groups. However the teacher / researcher who may be interested in exploring how students learn through multimedia and hypermedia and what behaviours and strategies can make them better learners could use tracking data and activity monitoring software.

8.5 Contribution

Ivory and Hearst (2001) make a case for improving the reliability, predictability and systematicity of MM usability evaluations by using several different techniques and increasing the number of participants of the evaluation. The scale of the evaluation undertaken for this study with reference to the sheer number of participants makes a contribution to the field of both usability evaluation and CALL evaluation.

This study explored the literature of five domains: (i) educational evaluation methods; (ii) educational research methods; (iii) HCI usability evaluation methods; (iv) ELT materials evaluation; (v) CALL evaluation. There is sufficient overlap and interdependency between (i) and (ii) to consider them both as belonging to one field. CALL evaluation borrows heavily from both fields and ELT materials evaluation borrows from (i) and (ii). Figure 8.1 represents graphically the positioning of this study within the field.

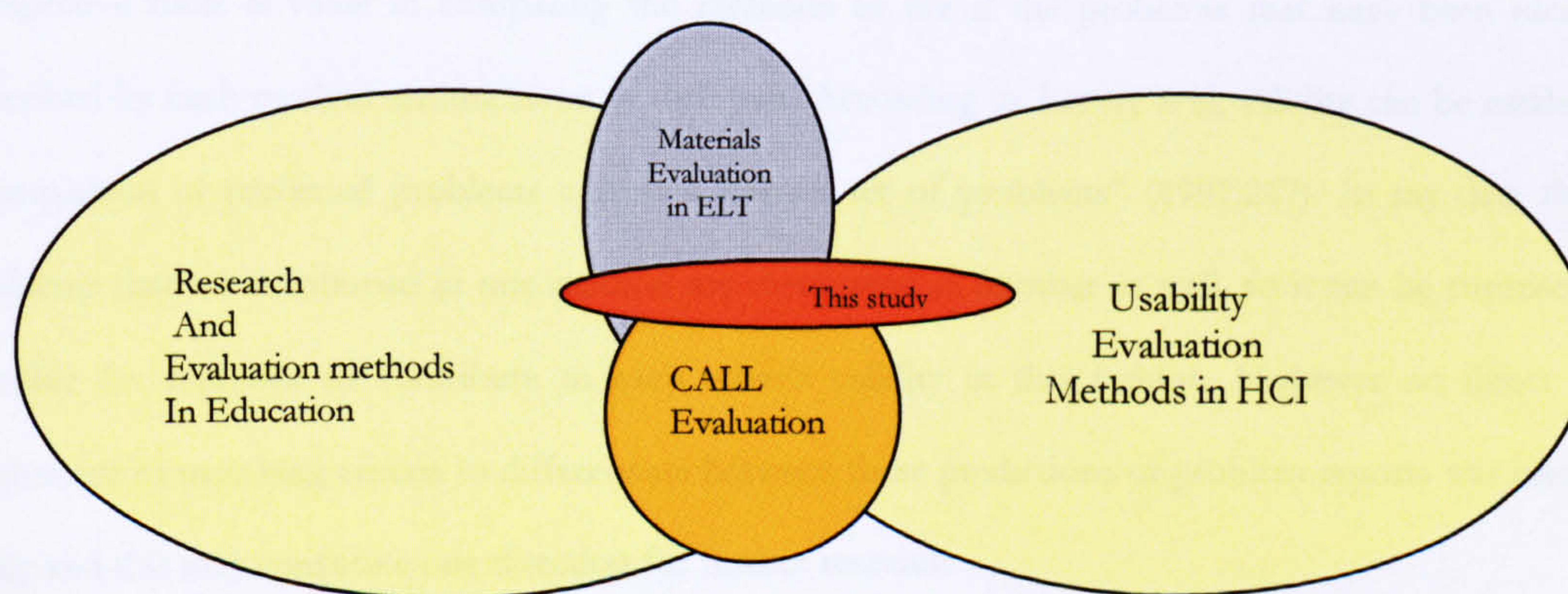


Figure 8.1: The positioning and contribution of this study

The study contributes to modern best practice in MM materials evaluation by investigating means of assessing the qualities and limitations of potential evaluation methods. The composite set of core criteria devised in this study and the way to use these criteria to gauge the potential of evaluation methods for CALL MM materials evaluation in ELT is a contribution. Moreover my commentary on the performance of the four methods on these core criteria is also a key contribution. The HCI literature takes into account the learning effectiveness of educational materials, and my evaluation methods incorporate insights from this literature. However I found that although there is a long tradition of paper based materials evaluation in ELT, there is a paucity of literature on ELT MM materials evaluation, amounting mainly to formative evaluations undertaken from the perspective of the materials developer rather than the end user. Very few studies in CALL and none in ELT MM investigate evaluation methods. Experts all stress the importance of evaluation but none have developed proposals for systematic best practice in the context of ELT MM, or explored the 'goodness of fit' between evaluation methods and evaluation purposes.

8.6 Suggestions for further research

This study investigated a variety of different methods before selecting four for closer assessment. It has attempted to assess the quality of the chosen methods by adopting the concerns of HCI evaluation methods. In HCI methods *validity*, *reliability* and *productivity* are important concerns of quality assessment along with *effectiveness* and *efficiency*. *Effectiveness* is measured through expert reviews or guideline-based analytic evaluation of results from empirical user-testing (Lavery *et al.* 1997). *Efficiency* is measured in terms of the incurred cost of analysis, and validity is measured in terms of whether the method suggests false problems (*ibid*). From the usability perspective there is value in comparing the methods to see if the problems that have been identified and described by each method are the same or different. According to Lavery *et al*, validity can be established by a "comparison of predicted problems against a known set of problems" (1997:247). In my data the usability problems that are mentioned in one method are mentioned in another as well, so it can be surmised that it is possible for methods to contribute to each other's validity in this fashion. However no direct measuring instrument or matching criteria to differentiate between these predictions or problem reports was created in this study and this may constitute one direction for further research.

Another area of further research is that of the timing of the evaluation. In chapter seven of this thesis a problem was identified with the double use of the SUMI usability questionnaire. Unexpectedly the results from the day one and day two evaluations (a week later) were different, although the respondents and the materials being evaluated were the same. This difference can be explained on the basis of the context of use for SUMI. On the first day the respondents were tired and at the end of a long day and an intensive course. On the second occasion the questionnaire was distributed in a morning session after they had been given the opportunity to put into practice all that they had learnt through the materials, thus establishing the efficacy of the materials for them personally. This finding suggests that the timing of an evaluation is crucial, not only in ELT MM materials evaluation, but also in educational evaluation generally. The HCI literature suggests that usability evaluations tend to be low on reliability (as discussed earlier), leaving open the question of whether the measuring instrument (e.g. SUMI) is at fault, or the general method (Questionnaires). Triangulation (through PLUM results) upheld the positive results of the second SUMI evaluation in this study, but the issue of *timing* was identified as a direction for future research.

The tracking data gathered from the activity monitoring also suggested another possible direction for future research. The data collected was so rich that it could be used to provide insights from many perspectives, other than simply navigation patterns and time usage.

Difficulties faced in preparing the tracking data from activity monitoring also highlighted the need for an automated data interpreting system. HCI methods are most definitely moving in the direction of automation of usability evaluation as it is expensive in time and human resources (Ivory and Hearst, 2001). CALL materials evaluation will need to follow suit.

Two recent papers Hémard (2006a and 2006 b) also emphasise the need to use HCI concepts to guide design and evaluation of CALL materials. Hémard emphasises the need for a strong HCI theoretical underpinning in the design and evaluation of CALL materials and the crucial importance of assessing “the performance and the process of engagement demonstrated by language learners when interacting within learning activities” (Hémard 2006a:270). His second study (2006b) takes this further and focuses on the interactivity of online materials, analysing the relationship between hypermedia/multimedia structures and language learning from an HCI perspective. This paper identifies the key features of user interface design and interactivity that affect the

learners' focus and level of engagement with the materials. H  mard suggests that the designer's model of a learning experience does not 'match' or 'fit' the learners' 'mental' model in online/hypermedia/distance learning. He says that while the hyper-linking of information is done easily with new authoring tools, little thought goes into how hyperlinking is used by the learner, whether the approach has been successful or not, and how feedback can be returned to the designer. A feature of H  mard's paper is the representation of key user interaction case scenarios using standardised Unified Modelling Language (UML) notation to show models and learner behaviours on mind-map/flowchart like diagrams⁴. (UML diagrams are quite similar to the diagrams used in my study).

H  mard (2006b) investigated the learners' use of materials by task analysis of a hypermedia prototype. This was combined with participatory methods such as user walkthroughs and focus groups to gather evaluation data. The objectivity of the prototype task analysis is triangulated with the learner-centred methods of focus groups and user walkthroughs, which elicit more subjective opinions. H  mard proposes a way to evaluate programs using UML, developing a loop that feeds user responses back into the design process, thus capturing learner activity with the system to enhance the learning program.

Automation of evaluation procedures with the aim of improving design is the ultimate ambition. However the limitation of this approach is that only the designer of the materials will be able to benefit from the automated procedure. Ivory and Hearst (2001) also discuss the automation of usability evaluation tools but it is unclear whether they envisage that the teacher evaluator should also be able to use such software, in contrast to walkthroughs and focus groups where no extensive training is required.

Nonetheless future research in CALL MM materials evaluation is likely to take the direction of a strong input from HCI theory with a move towards using more learner centric methods. We are also likely to see more automation in the field of usability evaluation.

⁴ UML is a language used by software developers for hypermedia representation in order to standardise object-oriented development.

8.7 Concluding remarks

This final chapter has tried to answer the research questions of the study and to make proposals for best practice in the evaluation of CALL MM materials. It has discussed the strengths and weaknesses of individual methods before presenting different combinations of the methods for different case scenarios. It has also indicated the contribution this study makes to the field, before identifying directions for further research. It is indeed gratifying to see that recent studies in the field (Hémard 2006a, 2006b) are also stressing the importance of applying principles from the domain of HCI to the design, development and evaluation of CALL MM materials, thus strengthening the recommendations of this study.

BIBLIOGRAPHY

- Ahmad, K., Corbett, G., Rogers, M., and Sussex, R (1985). *Computers, Language Learning, and Language Teaching*. Cambridge: Cambridge University Press.
- Alexander, S. and Hedberg, J.G. (1994). *Interactive Multimedia in University Education: Designing for Change in Teaching and Learning* (A-59) In Beattie, K., McNaught, C. and Wills, S. (eds) *Interactive Multimedia in University Education: Designing for Change in Teaching and Learning*. Elsevier Science B.V. (North-Holland).
- Allan, G. (2003). A critique of using grounded theory as a research method. *Journal of Business Research Methods* 2, (1)-10.
- Allum, P (2002). CALL and the classroom: the case for comparative research, *ReCALL* 14, 146-166
- Allwright, D. (1981). 'What do we want teaching materials for?' *ELT Journal* 36(1): 5-18 [reprinted in R. Rossner and R Bolitho (eds) (1990). *Currents of Change in English Language Teaching*, Oxford: Oxford University Press, 131-47].
- Allwright, R. (1982). 'Perceiving and pursuing learners; needs', in M. Geddes and G. Sturtridge (eds) (1982). *Individualization*, Oxford: Modern English Publications.
- Almstrum, V.L., Dale, N., Berglund, A., Granger, M., Little, J.C., Miller, D.M., Petre, M., Schragger, P., Springsteel, F. (1996) *Evaluation: turning technology from toy to tool*. Report of the Working Group on Evaluation. Integrating Tech. into CSE Barcelona, Spain.
- Altrichter, H. and Posch, P. and Somekh, B.(1996). *Teachers Investigate their Work. An Introduction to the Methods of Action Research*. London : Routledge.
- Atkinson, P. and Delamont, S. (1993) Bread and dreams or bread and circuses? A critique of 'case study' research in education. In M. Hammersley (ed.) *Controversies in Classroom Research. 2nd Edition*. Buckingham: Open University Press.
- BAAL(1994) *Recommendations on Good practice in Applied Linguistics* available at: http://www.baal.org.uk/about_goodpractice_full.pdf (accessed 11May 2006 10:30 a.m).
- Baddeley, A.D. (1976). *The Psychology of Memory*, New York, NY, Basic Books.
- Bader, M J (2000) Choosing CALL software: beginning the evaluation process, *TESOL journal* 9, 18-22
- Bassey, M (1981). Pedagogic research: On the relative merits of search for generalization and study of single events. *Oxford Review of Education*, 7(1), 73-94.
- Bassey, M (1999). *Case Study Research in Educational Settings*. Buckingham: Open University Press.
- Bates, A.W(1994) *Educational Multi-Media in a Networked Society* in Ottman, T. and Tomek, I. (1994) *Educational Multimedia and Hypermedia*, 1994 Charlottesville, Va.: Association for the Advancement of Computing in Education. [Available on line: <http://bates.cstudies.ubc.ca/edmedia.html>. Last accessed on 22/04/2006.also at: <http://www.tonybates.ca/papers/edmedia.html>
- Bates, A. (1997) 'The impact of technological change on open and distance learning', *Distance Education* 18 (1).

Bates, A.W. (1997) Technology, Distance Education and National Development. 18th ICDE World Conference [Online] available at: <http://bates.cstudies.ubc.ca/icde/icde.htm>. Last accessed on 22/04/2006.

Bayram,S and Nous, A. P.(2004).Evolution of educational software evaluation: instructional software assessment.*The Turkish Online Journal of Educational Technology – TOJET*. 3 (2), 4

Bax, S. (2003). "CALL-past, present and future". *System*, 31 (1),pp 13-28

Beatty.K (2003) *Teaching and Researching Computer Assisted Language learning*. UK Pearson Education Limited.

Beatty,K. and Nunan, D. (2004) Computer-mediated collaborative learning. *System*. 32 (2) Pages165-183

Bennett, J. (2003). *Evaluation Methods In Research*. London. Continuum

Bias. R(1994). "The Pluralistic Usability Walkthrough: Coordinated Empathies;, in J. Nielsen and R. Mack "Usability Inspection Methods", Chapter 3, pp.63-76, John Wiley.

Bloom, H. (1956) *Taxonomy of Educational Objectives, the Classification of Educational Goals – Handbook I: Cognitive Domain*. New York : McKay.

Bloor,M., Frankland, J., Thomas, M. and Robson, K. (2001) *Focus Groups in Social Research*. London: Sage

Boud, D., Keogh, R. & Walker, D. (eds.) (1985) *Reflection. Turning experience into learning*, London: Kogan Page.

Breen, M. and Candlin, C. (1987), 'Which materials? A consumer's and designer's guide', in L. Sheldon (ed.) (1987). *ELT Textbooks and Materials: Problems in Evaluation and Development*, ELT Documents 126, Oxford: Modern English Publications/the British Council, 13-28.

Brett, P.A. (1994).Using text reconstruction software. *English Language Teaching Journal* 48(4) 329-336.

Brett, P (2000) Integrating multimedia into the Business English curriculum: a casestudy. *English for specific purposes* 19, 269-290

Briggs, S. R., and Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54, 106-148.

Brown, A. (1981). Metacognition in reading and writing: The development and facilitation of selective attention strategies for learning from texts. In M. Kamil (Ed.), *Directions in reading research and instruction*. (pp. 21-43). Washington D.C. National Reading Conference.

Brown, A., Armbruster, B. and Baker, L. (1983). The role of metacognition in reading and studying . In J. Orsany (Ed.) *Reading comprehension: From research to practice*. (pp. 49-75). Hillsdale, NJ : Lawrence Erlbaun.

Brown. J.D. (1988) *Understanding Research in Second Language Learning: A teacher's guide to statistics and research design*. Cambridge. Cambridge University Press.

Bryman, A. (2004) *Social Research Methods*.(2nd Ed) New York. OUP.

Candlin, C. and Breen, M. (1979), *Evaluating and Designing Language Teaching Materials*, Practical Papers in English Language Education, vol. 2, University of Lancaster.

Carmines, E.G. & Zeller, R.A. (1979) *Reliability and Validity Assessment*, London: Sage Publications

Castellan, N. J. (1993). Evaluating information technology in teaching and learning. *Behavior Research Methods, Instruments, and Computers*, 25 (2), 233-237.

- Cauldwell, R. (2002) Streaming speech: listening and pronunciation for advanced learners of English. Birmingham: *Speech in Action*, 2002. Windows cd-rom
- Center for Educational Research and Innovation (CERI) (1989). Information Technologies in Education: The Quest for Quality Software. Organization for Economic Co-Operation and Development (OECD), Paris.
- Chambers, F. (1997). 'Seeking consensus in coursebook evaluation', *ELT Journal* 51(1): 29-35.
- Chignell, M. H., Motoyama, T. and Melo, V. (1995). Discount video analysis for usability engineering. In Anzai, Y., Ogawa, K. and Mori, H. (Eds). *Symbiosis of Human Artefact*, Volume 20B. Amsterdam: Elsevier Science B. V., 323-328.
- Churchill, G. (1991). *Marketing Research: Methodological Foundations*. London: The Dryden Press.
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53, 445-459.
- Clark, J.M. and Paivio, A. (1991) Dual Coding Theory and Education. *Educational Psychology Review*, Vol 3/3
- Cohen, A.D. (1998). *Strategies in learning and using a second language*. London. Longman
- Cohen, A and Hosenfeld, C., (1981). 'Some uses of mentalistic data in second language research.' *Language Learning* 31/2:285-313
- Cohen, L, Manion, L., and Morrison, K. (2000). *Research Methods in Education*. London: Routledge.
- Cohen, L., Manion, L., and Morrison, K. (2003) 5th Ed. *Research Methods in Education*. London, New York: Routledge, Falmer
- Collis, B. and Moonen, J. (2001) *Flexible Learning in a Digital World: Experiences and Expectations* London: Kogan Page
- Conyer, M. (1995). User and usability testing - how it should be undertaken? *Australian Journal of Educational Technology*, 11(2), 38-51. <http://www.ascilite.org.au/ajet/ajet11/conyer.html> Accessed on April, 12, 2006
- Corry, M. D., Frick, T. W. and Hansen, L. (1997). User-centered design and usability testing of a Web site: An illustrative case study. *Educational Technology Research and Development*, 45(4), 65-76.
- Creswell, J.W. (2003) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, (2nd Ed). London. Sage.
- Cronbach, L. J. (1951). *Coefficient alpha and the internal structure of tests*. Psychometrika. 16, 297-334.
- Cronbach, L. (1963) Course improvement through evaluation. *Teachers College Record*, 64, 672-83. Reprinted in M. Golby, J. Greenwald and R. West (eds) (1975) *Curriculum Design*. London: Croom Helm.
- Crowther, M., Keller, C., and G. Waddoups (2004). Improving the Quality and effectiveness of computer-mediated instruction through usability evaluations.' *British Journal of Educational Technology* 35(3): 289-303
- Cunningsworth, A. (1984), *Evaluating and Selecting EFL Teaching Materials*, London" Heinemann.
- Cunningsworth, A. (1995), *Choosing Your Coursebook*, Oxford: Heinemann.
- Daoud, A.-M. and Celce-Murcia, M. (1979), 'Selecting and evaluating a textbook', in Celce-Mucia, M. and McIntosh, L. (eds) (1979), *Teaching English as a Second or Foreign Language*, Rowley, MA: Newbury House, 302-7.

- Davies, G. and Higgins, J. (1985). *Using Computers in Language Learning: A Teacher's Guide* (2nd edn). London: Centre for information on Language Teaching Research.
- Davis, D.W. and Silver, B.D. (2003) Stereotype Threat and Race of Interviewer Effects in a Survey on Political Knowledge. *American Journal of Political Science* 47, 33-45.
- Delamont, S. (1978) Sociology in the classroom. In L. Barton and R. Meighan (eds) *Sociological Interpretations of Schools and Classrooms*. Driffield: Nafferton Books.
- Denzin, N.K. and Lincoln, Y.S. (2000) *Handbook of Qualitative Research*, (Eds. Denzin, Norman, K., Lincoln, Yvonna, S.), Sage Publications Ltd, London.
- Desmarais, L., Laurier, M. & Renié, D. (1998). The Analysis of Navigation Patterns in CALL. *Computer Assisted Language Learning* Vol. 11,(3,) pp. 309-315
- Dix, A., Finlay, J., Abowd, G., and Beale, R. (2004) 3rd ed. *Human-Computer Interaction*,. Pearson Education Limited. UK
- Dodson, C. (1967). *Language Teaching and the Bilingual Method*. London .Pitman
- Dublin, F. and Olshtain, E. (1986), *Course Design: Developing Programs and Materials for Language Learning*, Cambridge: Cambridge University Press.
- Dumas, J. S. and Redish, J. C. (1993). A practical guide to usability testing, Norwood, NJ: Ablex.
- Dunkel, P. (1991). The effectiveness of research on computer-assisted instruction and computer-assisted language learning. In P. Dunkel (ed.), *Computer-Assisted Language Learning and Testing: Research Issues and Practice* (pp.5-36). New York: Newbury House.
- Eastment, D. (1999) *The Internet and ELT*. Oxford. Summertown Publishing. Educational Technology and Society, 3(2)
- Ellis, R. (1998), 'The evaluation of communicative tasks', in B. Tomlinson (ed.) (1998a), *Materials Development in Language Teaching*, Cambridge: Cambridge University Press, 217-38.
- Ellis, R. (1997). The empirical evaluation of language teaching materials *English Language Teaching Journal* 51 (1) 36-42
- Ericsson, K. A. and Simon, H. A. (1985). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Ericsson, K.A. & Simon, H.A. (1987) 'Verbal Reports on Thinking', in C. Faerch & G. Kasper (eds.), *Introspection in Second Language Research*. Clevedon: Multilingual Matters, pp.9-45.
- Ericsson, K. A. and Simon, H. A. (1993), (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Færch, C. and Kasper, G. eds. (1987a). *Introspection in second language research*. Clevedon: Multilingual Matters.
- Færch, C. and Kasper, G. (1987b). "From product to process: Introspective methods in second language research". Færch and Kasper 1987a: 5-23.
- Felix, U. (2003a) Teaching Languages online: deconstructing the Myths, *Australian Journal of Educational Technology* AJET 19(1), 118-138 Accessed on 29/04/06
- Felix, U. (2003b) An orchestrated vision of language learning online. In U. Felix (Ed.) *Language Learning Online: Towards Best Practice*. Netherlands. Swets and Zeitlinger (pp.7-18)

Felix, U. (2003c) Pedagogy on the line: identifying and closing the missing links In U. Felix (Ed.) *Language Learning Online: Towards Best Practice*. Netherlands. Swets and Zeitlinger (pp.147-170).

Ferney, D., and Waller, S. (2000). Key factors in the Design of an Interactive Multimedia CD-Rom for EAP learning. In *EAP Learning Technologies*. Howarth, P., and Herington, R., (eds). Leeds University Press.

Flowerdew, L. (1994), 'Incorporating aspects of style and tone in self-access CALL courseware' in D. Gardner and L. Miller (eds) (1994), *Directions in Self-access Language Learning*, Hong Kong: Hong Kong University Press, 133-43.

Friend, D. (2005) Sky Pronunciation suite, Sky software house
<<http://www.skysoftwarehouse.com/programs.html>>

Gardiner, M.M and Christie, B (eds) (1987) *Applying Cognitive Psychology to User-Interface Design*. Chichester, John Wiley

Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.

Gibbs, A. (1997). 'Focus Groups', *Social Research Update*, Issue Nineteen. Department of Sociology. University of Surrey. Available at: <<http://www.soc.surrey.ac.uk/sru/SRU19.html>>

Glendinning, E and Howard, R. (2003), Lotus ScreenCam as an Aid to Investigating Student Writing, *Computer Assisted Language Learning*, 16, (1) 31 - 46, URL <<http://dx.doi.org/10.1076/call.16.1.31.15530>>

Glynn, C.J. and McLeod, J.M. (1984) Public Opinion du Jour: An Examination of the Spiral of Silence. *Public Opinion Quarterly* 48, 731-740.

Goldman, A. and MacDonald, S. (1987) *The group depth interview: Principles and practices*. Englewood Cliffs NJ: Prentice Hall.

Goodfellow, R. (1995). A review of the types of CALL programs for vocabulary instruction. *Computer-Assisted Language Learning*, 8(2-3), 205-226.

Goulding, C. (1998) Grounded Theory: the missing methodology on the interpretivist agenda. *Qualitative Marketing Research: An International Journal* 1, 50-57.

Goulding, C. (2002) Grounded Theory: A Practical Guide for Management, Business and Market Researchers. *Sage Publications Ltd, London*

Grant, N. (1987), *Making the Most of Your Textbook*, Harlow, Essex: Longman.

Greenbaum T L (1993) *The Handbook for Focus Group Research*. New York Lexington Books

Grellet, F. (1981). *Developing Reading Skills*, Cambridge: Cambridge University Press.

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass.

Guba, E. and Lincoln, Y. (1994) *Fourth Generation Evaluation*, Sage Publishing, London.

Gyselinck, V, M.-F Ehrlich, C. Cornoldi, R. de Beni and V. Dubois. (2000) Visuospatial working memory in learning from multimedia systems. *Journal of Computer Assisted Learning*, 16, 166-176.

Hair, J., Anderson, R., and Tatham, R. (2005). *Multivariate Data Analysis*, (4 ed.) Englewood Cliffs, NJ: Prentice-Hall.

- Hall, T. L. (1996). *Utilizing multimedia toolbox 3.0*, Danvers, MA: Boud and Fraser.
- Hammond (1989) Nick Hammond and Lesley Allinson. "Extending Hypertext for Learning: An Investigation of Access and Guidance Tools" in *Proceedings of the British Computer Society HCI 89*, 293-304, 1989.
- Hannafin, M. and Peck, K. (1988). *The Design, Development, and Evaluation of Instructional Software*. Macmillan Publishing, New York.
- Hargreaves, D. (1996) Teaching as a research-base profession: Possibilities and prospects. *Teacher Training Agency Annual Lecture*. London: The Teacher Training Agency (TTA).
- Harmer, J. (1983/1991), 'Appendix: Evaluating Materials', in J. Harmer, *The Practice of English Language Teaching*, Harlow, Essex: Longman, 276-84.
- Haugland, S. W. and Shade, D. D. (1990). Developmental evaluations of software for young children. Delmar Publishers Inc. Albany, New York.
- Hayes, T J. & Tatham, C.B. (1989) *Focus Group Interviews. A Reader*. 2nd Ed. Chicago, Ill: American Marketing Association.
- Hedberg, J., Harper, B. and Brown, C. (1993). Reducing cognitive load in multimedia navigation. *Australian Journal of Educational Technology (AJET)*, 9(2): 157-181
- Hémard, D. (1998) Knowledge representations in hypermedia CALL authoring: conception and evaluation. *Computer Assisted Language Learning*, 11(3):247-264
- Hémard, D. and Cushion, S. (2003a). Design and Evaluation of an Online Test: assessment conceived as a complementary CALL tool. *Computer Assisted Language Learning*, 16 (2-3):119-139
- Hémard, D. and Cushion, S. (2003b) Designing a CALL Package for Arabic While Learning the Language Ab Initio. *Computer Assisted Language Learning*, 16 (2-3), 259-266
- Hémard, D. (2003). Language Learning Online: designing towards user acceptability. In U. Felix (Ed.) *Language Learning Online: towards best practice*. Swets and Zeitlinger Publishers, 21-43.
- Hémard, D. (2004) Enhancing Online CALL Design: the case for evaluation. *ReCALL*, 16 (2): 502-519.
- Hémard, D. (2006a) Design Issues Related to the Evaluation of Learner – Computer Interaction in a Web-based Environment: Activities v. Tasks. *Computer Assisted Language Learning*. 19, (2 & 3), pp. 261 – 276
- Hémard, D. (2006b). Evaluating hypermedia structures as a means of improving language learning strategies and motivation. *ReCALL* 18 (1): 24-44.
- Higgins, J. and Johns, T. (1984). *Computers in Language Learning*. Reading, MA: Addison-Wesley.
- Higgins, J. (1995) *Computers in English Language Learning*. Oxford. Intellect
- Hofland, K., and Johansson, S., (1982) *Word frequencies in British and American English*. Bergen: The Norwegian Computing Centre for the Humanities.
- Hofstede, G. (1980). Motivation, leadership, and organization: Do American theories apply abroad. *Organizational Dynamics*, 9 (1), 42-63.
- Hofstede, G. (1983). The cultural relativity of organizational practices and theories. *Journal of International Business Studies*. 14 (1), 75-89.

- Hoogeveen, M. (1997) Towards a Theory of the Effectiveness of Multimedia Systems, *International Journal of Human Computer Interaction*, 9(2), 151-168.
- Horney, M.A. (1993) 'Case Studies of Navigational Patterns in Constructive Hypertext', *Computer and Education* 20: 67-82.
- Hubbard, P.L (1996) Elements of CALL methodology: Development, Evaluation, and Implementation. In Pennington, M (ed.), *The Power of CALL*. Houston, TX: Athelstan Publishers: 15-32
- Hubbard, P. (1988) An integrated framework for CALL courseware evaluation. *CALICO Journal*, 6 (2), 51-72.
- Hubbard, P. (1992). A methodological framework for CALL courseware development. In M. Pennington and V. Stevens (Eds.), *Computers in Applied Linguistics* (39-66). Philadelphia: Multilingual Matters.
- Hutchinson, T. (1987), 'What's underneath?: an interactive view of materials evaluation', in L. Sheldon (ed.) (1987), *ELT Textbooks and Materials: Problems in Evaluation and Development*, ELT Documents 126, Oxford: Modern English Publications/the British Council, 37-44.
- Hutchinson, T. and Torres, E. (1994), 'The textbook as agent of change', *ELT Journal* 48.4: 315-28 [reprinted in T. Hedge and N. Whitney (eds) (1996), *Power, Pedagogy and Practice*, Oxford: Oxford University Press, 307-23].
- Hutchinson, T. and Waters, A. (1987), *English for Specific Purposes: a Learning-Centred Approach*, Cambridge: Cambridge University Press.
- Hwu, F.(2003) Learners' Behaviors in Computer-Based Input Activities Elicited Through Tracking Technologies, *Computer Assisted Language Learning*, 16, (1), pp 5-29
- Hwu, F.(2007) .Learners' strategies with a grammar application: the influence of language ability and personality preferences'. *ReCALL*, 19: 21-38 Cambridge University Press
- ISO 13407/DIS (1997): "Human-centred design processes for interactive systems" International Standard.
- ISO 9241-10 (Ed), 1991, Ergonomic Requirements for Office Work with Visual Display Terminals, Part 10, *Dialogue Principles* Draft International Standard.
- ISO /DIS 9241-11(1995): Ergonomic Requirements for Office Work with Visual Display Terminals, Part 11, *Guidance on specifying and measuring usability*, Draft International Standard.
- ISO /DIS 9241-11(1999): Ergonomic Requirements for Office Work with Visual Display Terminals, Part 11, *Guidance on specifying and measuring usability*, International Standard.
- Ivory, M.Y & Hearst,M.A (2001)The State of the Art in Automating Usability Evaluation of User Interfaces.*ACM Computing Surveys*, Vol. 33, No. 4, December 2001.pp.470-516
- Jim'enez, M and P'erez, JM (2002) Learner autonomy and new technologies, *Educational Media International*, 39, (1) 61-68.
- Johnson, R.(1986) 'Selecting a coursebook: a realistic approach', in S. Holden(ed.) (1986), *Techniques of Teaching:From Theory to Practice*, Papers from the 1985 Bologna conference. Oxford. Modern English Publications/British Council 54-7
- Johnson, P. and Duberley, J. (2000) *Understanding Management Research*, Sage Publications Ltd., London.
- Johnson, J. C. and Weller, S. C. (2002) Elicitation techniques for interviewing. In J. F. Gubrium and J. A. Holstein (eds) (2002), *Handbook of Interview Research: Context and Method*, pp.491-514. Thousand Oaks: Sage.

- Karat, J. (1997) User-centred software evaluation methodologies, Handbook of HCI, 2nd edition, Elsevier Science, pp.689-704
- Kelly, T., Nesi, H. & R. Revell. (2001) *EASE: Volume One: Listening to Lectures*. CD-ROM. The University of Warwick.
- Kelly, T., Richards, L., & H. Nesi (2004) *EASE: Seminar Skills 1: Presentations*. CD-ROM. The university of Warwick.
- Kessler, G. and Plakans, L. (2001) Incorporating ESOL learners' feedback and usability testing in instructor developed CALL materials, *TESOL journal*, 10, 15-20
- Kiely, R. and Rea-Dickins, P. (2005). *Program Evaluation in Language Education*. Basingstoke: Palgrave Macmillan.
- Kim, J. and Mueller, C. W. (1978). Introduction to factor analysis: What it is and how to do it. Thousand Oaks, CA: Sage Publications, *Quantitative Applications in the Social Sciences Series*, No. 13.
- Kirakowski, J. and Corbett, M. (1993), SUMI: the Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24(3) 210-212
- Kirakowski, J. and Corbett, M. (1994), *SUMI: the Software Usability Measurement Inventory*. Human Factors Research Group, University College Cork, Ireland.
- Kirakowski, J. (1995) "The Software usability measurement inventory: background and usage", Human Factors Research Group, University College Cork, Ireland. In Jordan, P., Thomas, B. and Weerdmeester, B., (eds) *Usability Evaluation in Industry*, Taylor and Francis.
- Kirakowski, J. (1998) *SUMI User Handbook*. Human Factors Research Group, University College Cork, Ireland.
- Kitzinger J. (1994) 'The methodology of focus groups: the importance of interaction between research participants', *Sociology of Health* 16 (1): 103-21.
- Kitzinger J. (1995) 'Introducing focus groups', *British Medical Journal* 311: 299-302.
- Kolb, D. A. (1984) *Experiential Learning*, Englewood Cliffs, NJ.: Prentice Hall.
- Krueger R.A. (1988) *Focus groups: a practical guide for applied research*. London: Sage.
- Krueger R.A. (1994) *Focus Groups: a Practical Guide for Applied Research*. 2nd Ed. Thousand Oaks Sage.
- Krueger, R. A. & Casey, M. A. (2000). *Focus Groups: A Practical Guide for Applied Research*. Thousand Oaks, CA: Sage.
- Krueger, R. A., & Casey, M. (2001). *Focus groups* (Third ed.). California, USA: Sage Publications, Inc
- Lankshear A.J. (1993) 'The use of focus groups in a study of attitudes to student nurse assessment', *Journal of Advanced Nursing* 18: 1986-89.
- Lasagabaster, D. and Sierra, J.M. (2003) Students' evaluation of CALL software programs, *Educational Media International*, 40, 3 / 4, 293- 304
- Laurillard, D. (1993) *Rethinking University Teaching: a framework for the effective use of educational technology*. London; New York: Routledge.

- Laurillard, D. (1994). The role of formative evaluation in the process of multimedia. In: K. Beattie, C. McNaught, and S. Wills (Eds.), *Interactive multimedia in university education: Designing for change in teaching and learning*, Amsterdam: Elsevier, 287-293.
- Laurillard, D. (2003). Foreword. In U. Felix (Ed.) *Language learning online: Towards best practice*. Lisse: Swets & Zeitlinger Publishers. Pp. ix-x
- Lavery, D., Cockton, G. and Atkinson, M. P. (1997) Comparison of evaluation methods using structured usability problem reports. *Behaviour and Information Technology*, VOL. 16, NO. 4/5, 246 ± 266
- Lawton, D. (1980) *The Politics of the School Curriculum*. London: Routledge and Kegan Paul.
- Lawton, D. (1983) *Curriculum Studies and Educational Planning*. London: Hodder and Stoughton.
- Lee, S.H (1999) 'Usability Testing for Developing Effective Interactive Multimedia Software: Concepts, Dimensions, and Procedures'. *Educational Technology and Society* 2(2)
- Lee, J. (1975) 'Choosing and using a textbook', *English Teaching Forum* 13:3/4 (Special Issue Pt 2): 364-4.
- Legenhausen L., and Wolff, D. (1990) CALL in use-use of CALL: Evaluating CALL software. *System*, Vol. 18. No. 1. pp. I-13.
- Levy, M. (1997). *Computer-Assisted Language Learning: Context and Conceptualisation*. Oxford: Oxford University Press.
- Lincoln, Y. and Guba, E. (1985) *Naturalistic Inquiry*. London: Sage Publications.
- Lincoln, Y.S. and Guba, E. (2000) 'Paradigmatic controversies, contradictions and emerging confluences' in *Handbook of Qualitative Research*, Sage Publishing Ltd., London.
- Litoselliti, L. *Using Focus Groups in Research*, Continuum: London, (2003)
- Liu, M., Ayersman, D. and Reed, W. (1995). Perceptions of hypermedia environment.' *Computers in Human Behaviour* 11/3-4: 411-428
- Low, G. (1987), 'The need for a multi-perspective approach to the evaluation of foreign language teaching materials', *Evaluation and Research in Education* 1.1: 19-29.
- Low, G. (1989), 'Appropriate design: the internal organisation of course units', in R. K. Johnson (ed.) (1989b), *The Second Language Curriculum*, Cambridge: Cambridge University Press, 136-54.
- Low, G.D. (1991). Talking to questionnaires: pragmatic models in questionnaire design. In P. Adams, B. Heaton & P. Howarth (Eds.), *Socio-Cultural Issues in English for Academic Purposes*, pp 118-143. London: MEP & The British Council.
- Lynch, B.K. (1990) A context adaptive model for program evaluation. *TESOL Quarterly*, 24(1), 23-42
- Lynch, B.K. (1996) *Language program evaluation*. Cambridge: Cambridge University Press
- Lynch, B.K. (2003) *Language Assessment and Programme Evaluation*. Edinburgh Textbooks in Applied Linguistics. Edinburgh. Edinburgh University Press.
- Lynch, T. (1996), 'Influences on course revision: an EAP case study', in M. Hewings and T. Dudley-Evans (eds) (1996), *Course Evaluation and Design in EAP*, Review of ELT 6.1, Hemel Hempstead, Herts: Prentice Hall Macmillan, 26-35.

- MacDonald, B.** (1976) Evaluation and the control of education. In D. Tawney (ed.) *Curriculum Evaluation Today: Trends and Implications*. London: MacMillan.
- Maley, A.** (1998). 'Squaring the circle - reconciling materials as constraint with materials as empowerment,' in B. Tomlinson (ed.) (1998) *Materials Development in language Teaching*, Cambridge: Cambridge University Press, 279-294.
- Malhotra, N.K. and Birks, D.** (2006) *Marketing Research, An Applied Orientation*, 3rd European Edition, Prentice-Hall.
- Mann, S.** (1983) 'Verbal reports as data: A focus on retrospection' in S. Dingwall, S. Mann, and F. Katamba (eds.): *Problems, Methods and Theory in Applied Linguistics*. Lancaster: Lancaster University.
- Mariampolski, H.** (2001) *Qualitative Market Research; A Comprehensive Guide*. Sage Publications, London
- Marmolin, H.** (1991). Multimedia from the Perspectives of Psychology. In Kjeldahl, L. (Ed.). *Multimedia. Systems, Interactions and Applications. 1st Eurographics Workshop, Stockholm, Sweden April 18-19, 1991*. Berlin, Germany: Springer-Verlag.
- Martin, D.** (2000) 2nd Ed. *Manipulating Meetings: How to get what you want when you want it*. Trans-Atlantic Pubns.
- Masuhara, H.** (1998). 'What do teachers really want from coursebooks?', in B. Tomlinson (ed.) (1998), *Materials Development in Language Teaching*, Cambridge: Cambridge University Press, 239-60.
- Matthews, A.** (1991). 'Choosing the best available textbook.' in Matthews, A. M. Spratt and L. Dangerfield (ed.) (1991), *At the Chalkface*. London: Nelson, 202-6.
- McDonough, J. and Shaw, C.** (1993). *Materials and methods in ELT : a teacher's guide* Oxford : Blackwell,
- McGrath, I.** (2002) *Materials Evaluation and Design for Language Teaching*. Edinburgh Textbooks in Applied Linguistics. Edinburgh University Press
- McKerlie, D. and Preece, J.** (1993). The hype and the media: Issues concerned with designing hypermedia. *Journal of Microcomputer Applications*, 16, 33-47.
- Mertens, D. M.** (2005) *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods*. 2nd edn. Thousand Oaks, California. Sage Publications
- Merton R.K., and Kendall P.L.** (1946) 'The Focused Interview', *American Journal of Sociology* 51: 541-557.
- Merton, R. K., Fiske, M., & Kendall, P.L.** (1990) *The focussed interview* (2nd Ed) New York : Free Press
- Miles, M. and Huberman, M.** (1994) *Qualitative Data Analysis: An Expanded Sourcebook*. 2nd edn. London: Sage.
- Molich, R., Bevan, N., Curson, I., Butler, S., Kindlund, E., Miller, D., & Kirakowski, J.** (1998). Comparative evaluation of usability tests. *Proceedings of Usability Professionals Association annual Conference, 1998* Washington DC.
- Morgan D.L.** (1988) *Focus groups as qualitative research*. London: Sage.
- Morgan, D.L.** (ed.) (1993) *Successful Focus Groups: Advancing the State of the Art*. Newbury Park, CA: Sage
- Morgan D.L. and Kreuger R.A.** (1993) 'When to use focus groups and why' in Morgan D.L. (Ed.) *Successful Focus Groups*. London: Sage.
- Morgan, D.L.** (1997) *Focus Groups as Qualitative Research*. 2nd edition. London, Thousand Oaks CA: Sage

- Murphy, D., Walker, R. and Webb, G. (eds) (2001) *Online Learning and Teaching with Technology: Case studies, Experience and Practice* London: Kogan Page
- Mydlarski, D. and Paramskas, D. (1991) *Vi-Conte*. Calgary: University of Calgary/PICS.
- Nardi, B. A. (ed.) (1996) *Context and Consciousness: Activity Theory and Human-Computer Interaction*. Cambridge, Massachusetts; London, England: MIT Press.
- Nesi, H. (1996) Roles for trainers and trainees in Computer-mediated courses. *ELTED* 2. (1): 1-15.
- Nielsen, J., and Molich, R. (1990). Heuristic evaluation of user interfaces, *Proceedings ACM CHI'90 Conference*. (Seattle, WA, 1-5 April), 249-256.
- Nielsen, J. (1993) *Usability Engineering*, Academic Press. Cambridge, MA: AP Professional.
- Nielsen, J. & Mack, R.L.(eds) (1994) *Usability Inspection Methods*. John Wiley & Sons Inc.
- Nielsen, J. (1994) Heuristic evaluation. In Nielsen, J. & Mack, R.L.(eds) *Usability Inspection Methods*. Chapter 2 pp25-61 .John Wiley .
- Nielsen, J. (1995) *Multimedia and Hypertext. The Internet and Beyond*. NY; London: AP Professional.
- Norris, N. (1990) *Understanding Educational Evaluation*. London: Kogan Page.
- Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw Hill.
- Nunan, D. (1992), *Research Methods in Language Learning*. Cambridge. CUP
- Oakley, A. (2000) *Experiments in Knowing*. Cambridge: Polity Press.
- Olshavsky, J.E. (1977) Reading as Problem Solving: An Investigation of Strategies. *Reading Research Quarterly*, 2, (4): 654-674.
- Orey, M.A. & Nelson, W.A. (1994) 'Visualizing Techniques for Examining Learner Interactions with Hypermedia Environments' (ERIC Document Reproduction Service No. ED 373 747).
- Owston, R.D. and Dudley-Marling, C. (1988). A criterion-based approach to software evaluation. *Journal of Research in Computing in Education*, Vol 20(3).
- Parlett, M. and Hamilton, D. (1976). Evaluation as illumination: A new approach to the study of innovative programmes. in D. Tawney (ed.) *Curriculum Evaluation Today: Trends and Implications*. London: Macmillan. Originally printed in (1972) *Occasional Paper No. 9*, Edinburgh: Centre for Research in the Educational Sciences, University of Edinburgh.
- Paivio, A. (1971). *Imagery and Verbal Processes*, Holt, Rinehart, and Winston, New York (Reprinted 1979, Erlbaum, Hillsdale, New Jersey).
- Paivio, A. (1986). *Mental Representations: A Dual-Coding Approach*, Oxford University Press, New York.
- Peirce, W. (2000). Online strategies for teaching thinking. *Syllabus*, 14(2), pp. 21, 24.
- Pérez, J.M. (2000). Learner autonomy and ICT: a web-based course for psychology, *Educational Media International*, 37, 257-261.
- Pica, T. (1994). Research on negotiation: what does it tell us about second-language learning conditions, processes and outcomes? *Language Learning*, 44, 493-527

PLUM Pre-Use Post use Questionnaires Programme on Learner Use of Media (PLUM) Institute of Educational Technology, The Open University, Milton Keynes Available at <http://iet.open.ac.uk/evaluation/PreProgramQs.html> & <http://iet.open.ac.uk/evaluation/PostProgramQs.html>

Last accessed on 16/2/2007

PLUM (the Programme on Learner Use of Media), Institute of Educational Technology, Open University of Educational Technology, UK, and **TELL** (Technology Enhanced Language Learning, a UK-TLTP consortium, led by the University of Hull).

Polson, P.G., Lewis, C., Rieman, J. & Wharton, C. (1992) 'Cognitive Walkthroughs: A Method for Theory-Based Evaluation of User Interfaces', *International Journal of Man-Machine Studies* 36: 741-73.

Porteous, M., Kirakowski, J., Corbett, M. (1993). *SUMI user handbook*. Human Factors Research Group, University College Cork, Ireland.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., and Carey, T. (1994). *Human-Computer Interaction*, Workingham, England: Addison-Wesley.

Puchta, C., and Potter, J. (2004), *Focus Group Practice*. London: Sage

Quentin-Baxter .M (1999). Quantitative Evidence for Differences Between Learners Making Use of Passive Hypermedia Learning Environments. *ACM Computing Surveys* 31(4)

Qui, L. (1994) 'Frequency Distribution of Hypertext Path Patterns: A Pragmatic Approach', *Information Processing and Management* 30: 131-40.

Rea-Dickins, P. (1994), 'Evaluation and English language teaching. State-of-the-art article', *Language Teaching* 27: 71-91.

Rea-Dickins, P. and Germaine, K. (1998), The price of everything and the value of nothing: trends in language program evaluation. In P. Rea-Dickins & K.P Germaine, (Eds.). *Managing evaluation and innovation in language teaching: Building Bridges*. London. Longman

Recker, M.M. (1994) 'A Methodology for Analysing Students' Interactions within Educational Hypertext' *ED-MEDIA, Educational Multimedia and Hypermedia Annual*, Vancouver.

Reiser, R. A., & Dick, W. (1990). *Evaluating instructional software*. Educational

Reiser, R. A., and Kegelmann, H. W. (1994). Evaluating instructional software: A review and critique of current methods. *Educational Technology Research and Development*, 42 (3), 63-69.

Reiterer, H. and Opperman, R. (1995). Standards and software-ergonomic evaluation. In Anzai, Y., Ogawa, K. and Mori, H. (Eds). *Symbiosis of Human Artefact*, Volume 20 B. Amsterdam: Elsevier Science B. V., 361-366.

Richards, J. (1998), 'Textbooks: help or hindrance in teaching?', Ch. 7, in J. Richards, *Beyond Training: Perspectives on Language Teacher Education*, Cambridge: Cambridge University Press, 125-40.

Richards, J. and Rodgers, T. (1986), *Approaches and Methods in Language Teaching*, Cambridge: Cambridge University Press.

Richards, J. and T. Rogers. (1982) Method: Approach, design, and procedure. *TESOL Quarterly*, 16 (2), 153-168.

Richards, K. (2003) *Qualitative Inquiry in TESOL*. New York, NY. Palgrave-Macmillan

- Roberts.C (1998) Language and cultural issues in innovation: The European dimension. In P. Rea-Dickins & K.P Germaine, (Eds.). *Managing evaluation and innovation in language teaching: Building Bridges*. London. Longman
- Roberts, J. T.(1996) Demystifying materials evaluation. *System* 24 (3) 375-390
- Ross.S.J (2003) A Diachronic Coherence Model for Language Program Evaluation *Language Learning* 53:1. March pp.1-33
- Rosson, M.B. & Carroll, J.M (2002) *Usability engineering : scenario-based development of human-computer interaction*. San Francisco ; London : Morgan Kaufmann
- Russel,J.D. & Blake,B.L. (1988). Formative and summative evaluation of instructional products and learners. *Educational Technology*, 28(9),22-28.
- Salaberry, R.(1996).A theoretical foundation for the development of pedagogical tasks in computer mediated communication. *CALICO journal*,14 (1): 5-34.
- Scantlebury, K. C. ; Boone, W. J. ; and Butler, K. J. (2001). Design, validation, and use of an evaluation instrument for monitoring systemic reform. *Journal of Research in Science Teaching*. 38(6), 646-662.
- Scarborough, D. (1988). 'Survey review: software for English language teaching.' *ELT Journal* 42(4):301-14.
- Scholfield, P and Ypsiladis, G (1994) Evaluating computer assisted language learning from the learners' point of view. In Graddol, D and Swann, J (eds) *Evaluating Language: Papers from the Annual Meeting of the British Association for Applied Linguistics Held at the University of Essex, September 1992*, Multilingual Matters,Clevedon.
- Schopmeyer, K.D and Fisher,B. J. (1993)Insiders and Outsiders: Exploring Ethnocentrism and Cultural Relativity in Sociology Courses. *Teaching Sociology*, Vol. 21, No. 2 pp. 148-153
- Schwier, R.A. and Misanchuk, E.R. (1993) *Interactive Multimedia Instruction*. Englewood Cliffs, NJ: Educational Technology Publications.
- Scriven, M. (1973) Goal-free evaluation. In E. House (ed.) *School Evaluation: The Politics and Process*. Berkeley, California: McCutchan Publishing Corporation.
- Scriven, M. (1990). The evaluation of hardware and software, *Studies in Educational Evaluation*, 16, 3-40.
- Seliger, H.W. and Shohamy, E. (1989) *Second Language Research Methods*. Oxford: OUP
- Shackel, B. (1991). Usability-Context, framework, definition, design and evaluation. In: B. Shackel and S. J. Richardson (Eds.), *Human Factors for Informatics Usability*, Cambridge: Cambridge University, 21-37.
- Shaw, I.F. (1999) *Qualitative Evaluation*. London:Sage
- Sheldon, L. (1988), 'Evaluating ELT textbooks and materials', *ELT Journal* 42.4: 237-46.
- Silverman, D. (2001) *Interpreting Qualitative Data - Methods for Anahysing Talk, Text and Interaction*, Sage Publications, London.
- Skierso, A. (1991), 'Textbook selection and evaluation', in M. Celce-Murcia, (ed.) (1991) (2nd edn), *Teaching English as a Second or Foreign Language*, Boston: Heinle and Heinle, 432-453.
- Smidt, E. & Hegelheimer,V. (2004) Effects of Online Academic Lectures On ESL listening Comprrehension, Incidental Vocabulary Acquisition, and Stategy Use. *Computer Assisted Language Learning. Vol* 17(5) pp517-556

- Smith, S. L. and Mosier, J. N. (1986). *Guidelines for Designing User Interface Software*. Report ESD-TR-86-278, The MITRE Corporation, Bedford MA 01730. (Available from National Technical Information Service, Springfield, VA.)
- Smith, M. and Glass, G. (1987) *Research and Evaluation in Education and the Social Sciences*. Englewood Cliffs, New Jersey: Prentice Hall.
- Spellman Miller, K., (1996) "Different tasks, different processes? An investigation of revision behavior on two academic tasks". *Conference Proceedings Second Language Acquisition and writing: A multidisciplinary approach*. University of Southampton, July 11-12.
- Spellman Miller, K., (2002) "Units of production in writing: evidence of topic 'framing' in on line writing research". *Reading Working Papers in Linguistics*. Vol:6. 255-272
- Squires, D., and McDougall, A. (1995). A critical examination of the checklist approach in software selection. *Journal of Educational Computing Research*, 12 (3), 263-274.
- Squires, D., and McDougall, A. (1996). Software evaluation: A situated approach. *Journal of Computer Assisted Learning*, 12, (3), 146-161.
- Squires, D. and Preece, J. (1999). Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. *Interacting with Computers* Volume 11, Issue 5, Pages 467-483.
- Stake, R. (1967) The countenance of educational evaluations. *Teachers College Record*, 68, 523-40.
- Stake, R. (1986) Evaluating educational programmes. In D. Hopkins (ed.) *Inservice Training and Educational Development*. London: Croom Helm.
- Stenhouse, L. (1975) *An Introduction to Curriculum Research and Development*. Oxford: Heinemann Educational Books.
- Stepp_Greany, J. (2002) 'Student perceptions on language learning in a technological environment: implications for the new millennium.' *Language Learning & Technology* 6/1: 165-180
- Stevens, V. (1996). *Software Selection, Evaluation and Use* Prepared by Vance Stevens, Courseware Publishing International, Cupertino, CA An e-mail communication sent to Vance Stevens from Deborah Healey, OSU/ELI July 1996, regarding how she presents Hubbard's Framework for CALL Evaluation presented at the 1996 Summer Technology Seminars held August 12-15 at the English Language Institute, Oregon State University, Corvallis, Oregon. Available at http://oregonstate.edu/Dept/eli/vstevens/deb_hubb.htm last accessed on 2006-12-31.
- Stevick, E. (1972), 'Evaluating and adapting language materials', in H. Allen, and R. Campbell (eds) (1972), *Teaching English as a Second Language: A Book of Readings*, New York: McGraw Hill, 101-20.
- Stewart D.W. and Shamdasani P.N. (1992) *Focus groups: theory and practice*. London: Sage.
- Sykes, W. (1990), 'Issues in qualitative research', *Journal of the Market Research Society*. Vol 33(1). pp3-12
- TESOL CALL IS Software List <http://oregonstate.edu/dept/eli/softlist/pub99.html>. Last updated 03/23/2006 by Deborah Healey. Accessed on 23 March 2006.
- Tierney, W.G. and Dilley, D. (2002) Interviewing in Education In J. F. Gubrium and J. A. Holstein (eds) (2002), *Handbook of Interview Research: Context and Method*, pp.453-471. Thousand Oaks: Sage.
- Tolhurst, D. (1995). Hypertext, hypermedia, multimedia defined? *Educational Technology*, 35(2), 21-26.

- Tucker, C.A., (1975). Evaluating beginning coursebooks. *English Language Teaching Forum* XIII/3/4, pp. 355–361.
- Tucker, C.A. (1978), 'Evaluating Beginning textbooks' reprinted as Appendix 3 in H. Madsen and J Bowen(1978), *Adaptation in Language Teaching*, Rowley, MA: Newbury House, 219-37. [First printed in 1975 in *English Teaching Forum* 13: (3-4)(Special issue pt-2) 355-61[originally printed in 1968 in *English Teaching Forum*, 6, 5 pp. 8-15, and subsequently reprinted as Appendix 3 in H. Madsen and J Bowen (1978).
- Ur, P. (1996), *A Course in Language Teaching*, Cambridge: Cambridge University Press.
- Vaughan,T. (1993) *Multimedia: Making it Work*. Berkeley,Calif. ;London: Osborne McGraw Hill.
- Warren, E. (1997). 'The significance of pauses in written discourse: a comparison of native speaker and learner writing.' In Archibald,A. & Jeffery,G.C. (eds) *Second Language Acquisition and Writing: A multi-disciplinary approach*. Conference 11th -12th July,1996 Proceedings. University of Southampton. Pp 152-177.
- Watkins, C., Carnell, E., Lodge, C., Wagner, P., and Whalley, C. (2000) *Learning about Learning: resources for promoting effective learning*, London: Routledge
- Watts,N. (1997). 'A learner-based design model for interactive multimedia language learning packages.' *System* 25 (1):1-8
- Weir, C. and Roberts, J. (1994). *Evaluation in ELT*: Oxford: Blackwell
- Wharton C, Rieman,J, .Lewis.C, and Polson. P(1994). *The Cognitive Walkthrough: A practitioner's guide*. In Jakob Nielsen and Robert L. Mack, editors, *Usability Inspection Methods*. John Wiley and Sons, Inc.
- Whitefield, A., Wilson, F., & Dowell, J. (1991). A framework for human factors evaluation. *Behaviour and Information Technology* 10, 1, 65–79.
- Whitelock, D. (2000). 'Frameworks for the Evaluation of Multimedia Learning Technologies:Lessons learned and future directions?' *Programme on Learner Use of Media (PLUM) Paper No. 131* Institute of Educational Technology, The Open University, Milton Keynes
- Wilkinson, S. (1998) 'Focus Groups—A feminist method', *Psychology of Women Quarterly* 23(2): 221-44.
- Williams, D. (1983). 'Developing criteria for textbook evaluation', *ELT Journal* 37(3), 251-5.
- Williams, M.D. and Dodge, B.J. (1993). *Tracking and analysing learner-computer interaction*. In Simonson, M.R. & Abu-Omar, K. (eds) Annual proceedings of selected research and development presentations at the 1993 national convention of the Association for Educational communications and Technology New Orleans, pp. 1115-1129.
- Wilson, H. W. (2000). How to evaluate educational software, *Curriculum Review*, 39(1), 15.
- Wilson, H. W. (2001). Beyond data mania, *Leadership*, 31(2), 8-36.
- Wilson, N. and McLean, S. (1994). *Questionnaire Design: a Practical Introduction*. Newton Abbey, Co. Antrim. University of Ulster Press.
- Yildiz, R and Atkins, M (1993) Evaluating Multimedia Applications, *Computers and Education*, 21, 133-139
- Yin,R.(1984). *Case Study Research*. Beverly hills, Calif.: Sage Publications

Yoder E.A., McCracken.D.L, and R. M. Akscyn,(1988)KMS: A distributed hypermedia system for managing knowledge in organizations. *Communications of the ACM*, Vol 31, Issue 7 ,pp. 820-835

APPENDICES

APPENDIX 1.1

APPENDIX 1.1: DESCRIPTION AND APPEARANCE OF MATERIALS

Description of EASE Listening to Lectures

Organisation Chart of Listening to Lectures

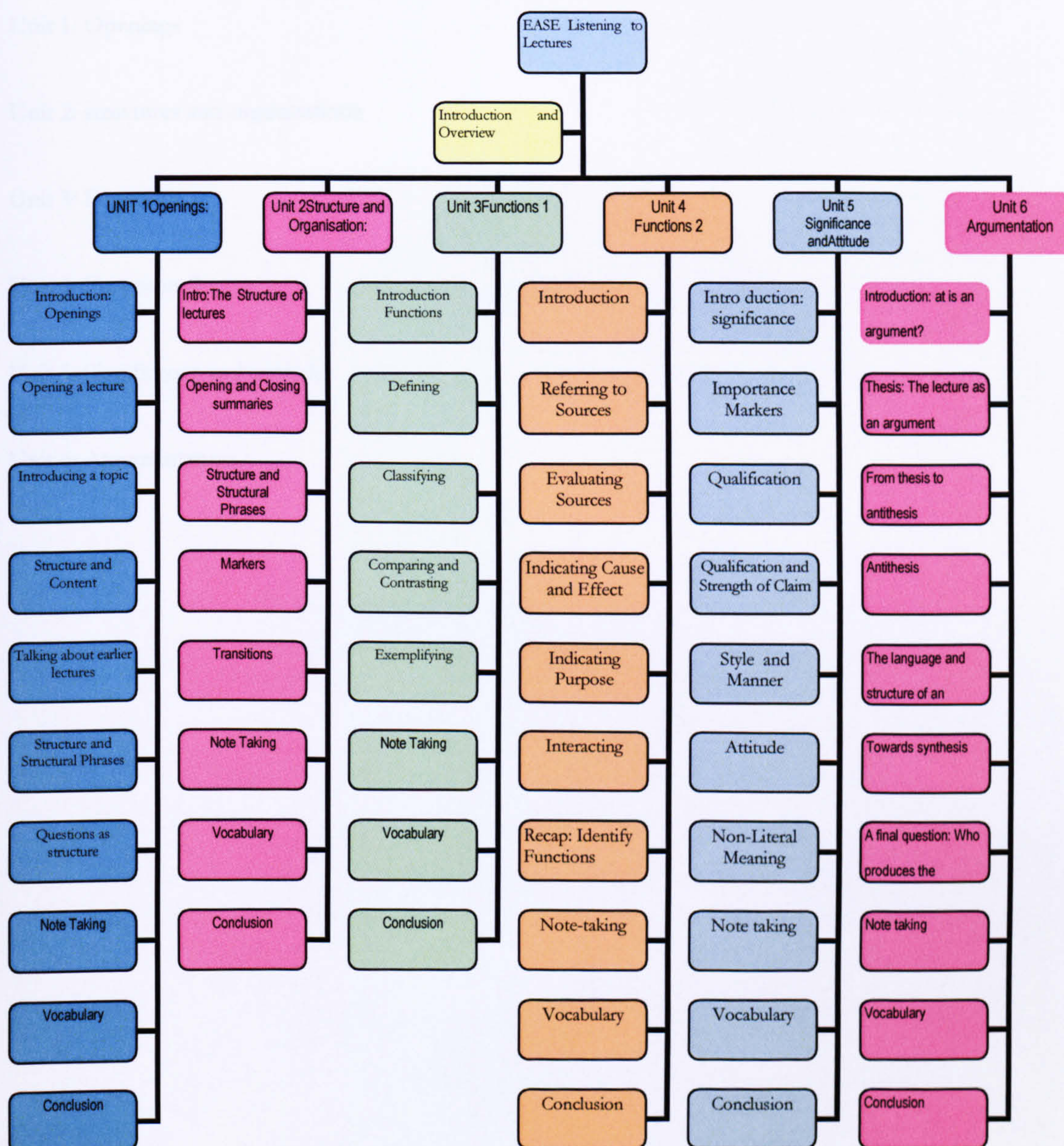


Figure: 1.1 Organisation chart of Listening to Lectures

When the *EASE Listening to Lectures* is opened at the start the program gives us options to select one of the seven units. These units are:

Introduction and overview

Unit 1: Openings

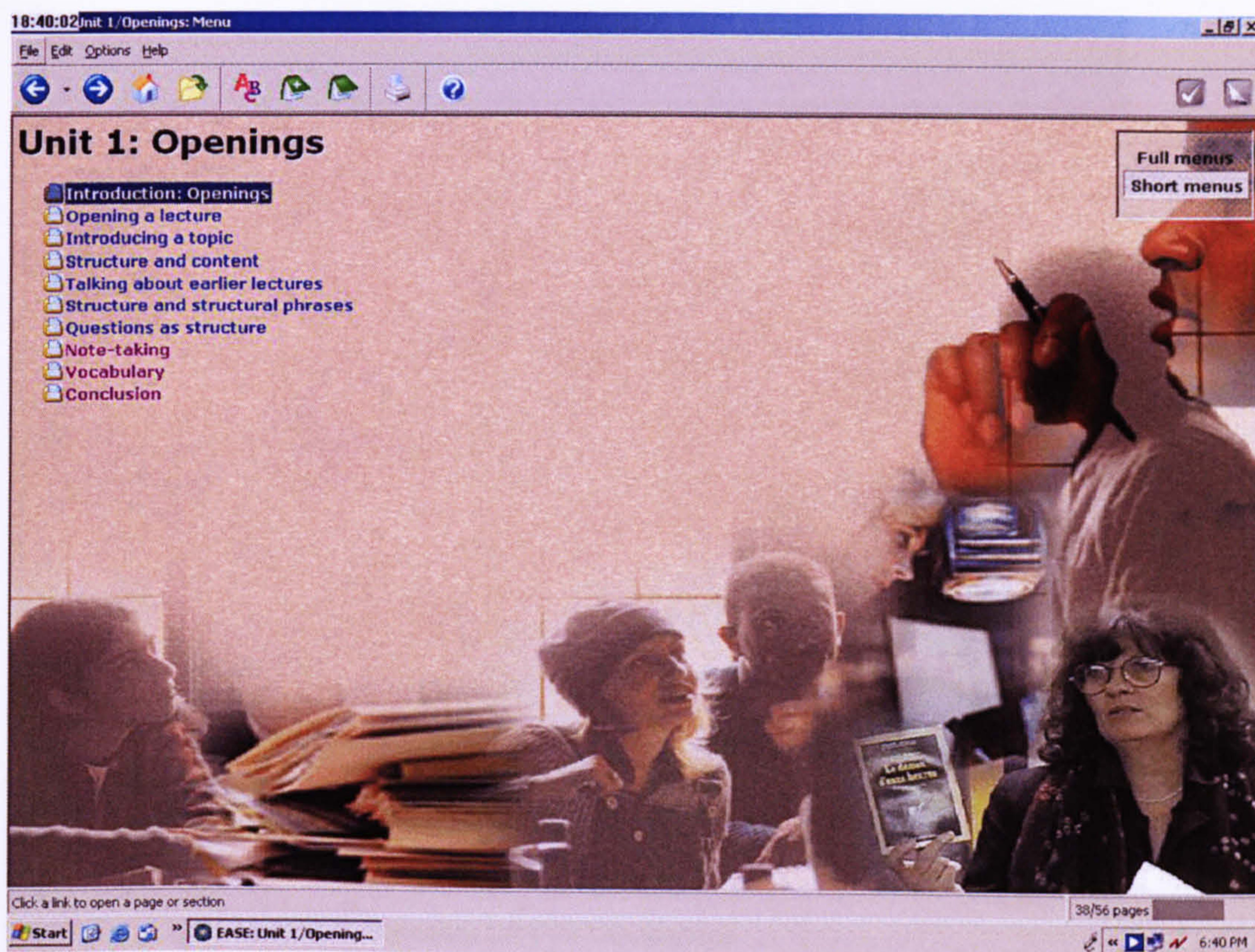
Unit 2: structures and organizations

Unit 3: Functions 1

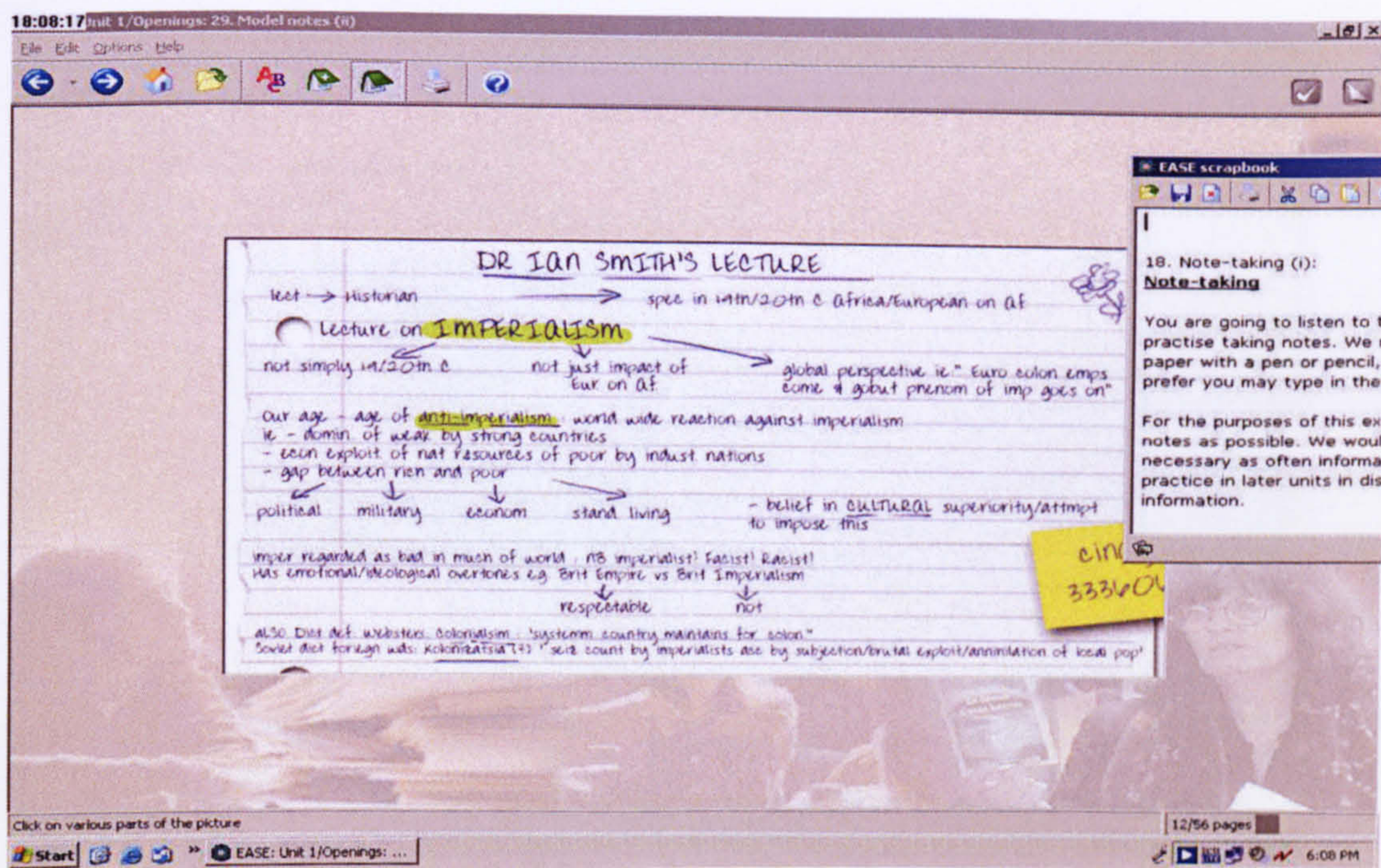
Unit 4: Functions 2

Unit 5: Significance and attitude

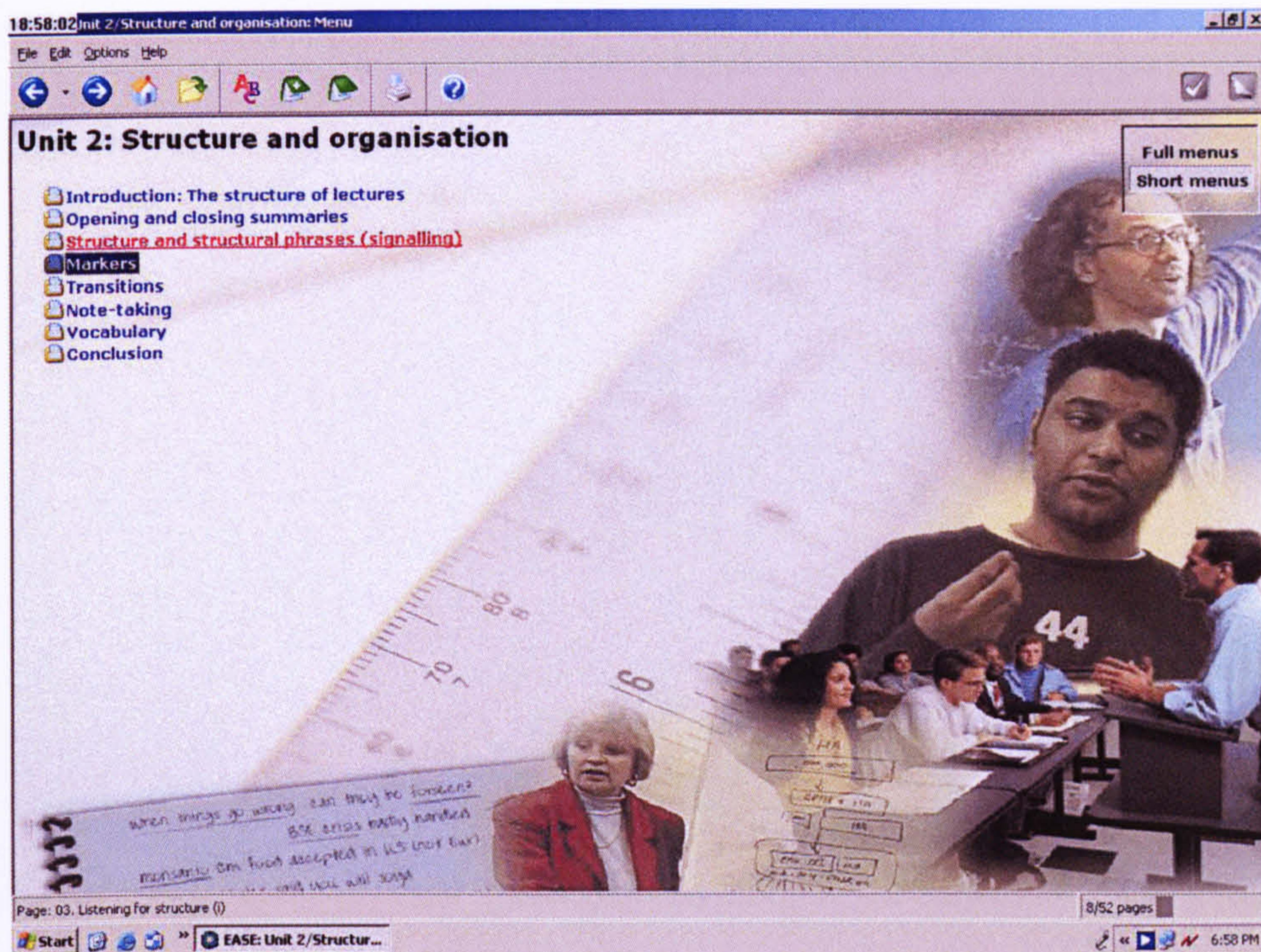
Unit 6: Argumentation.



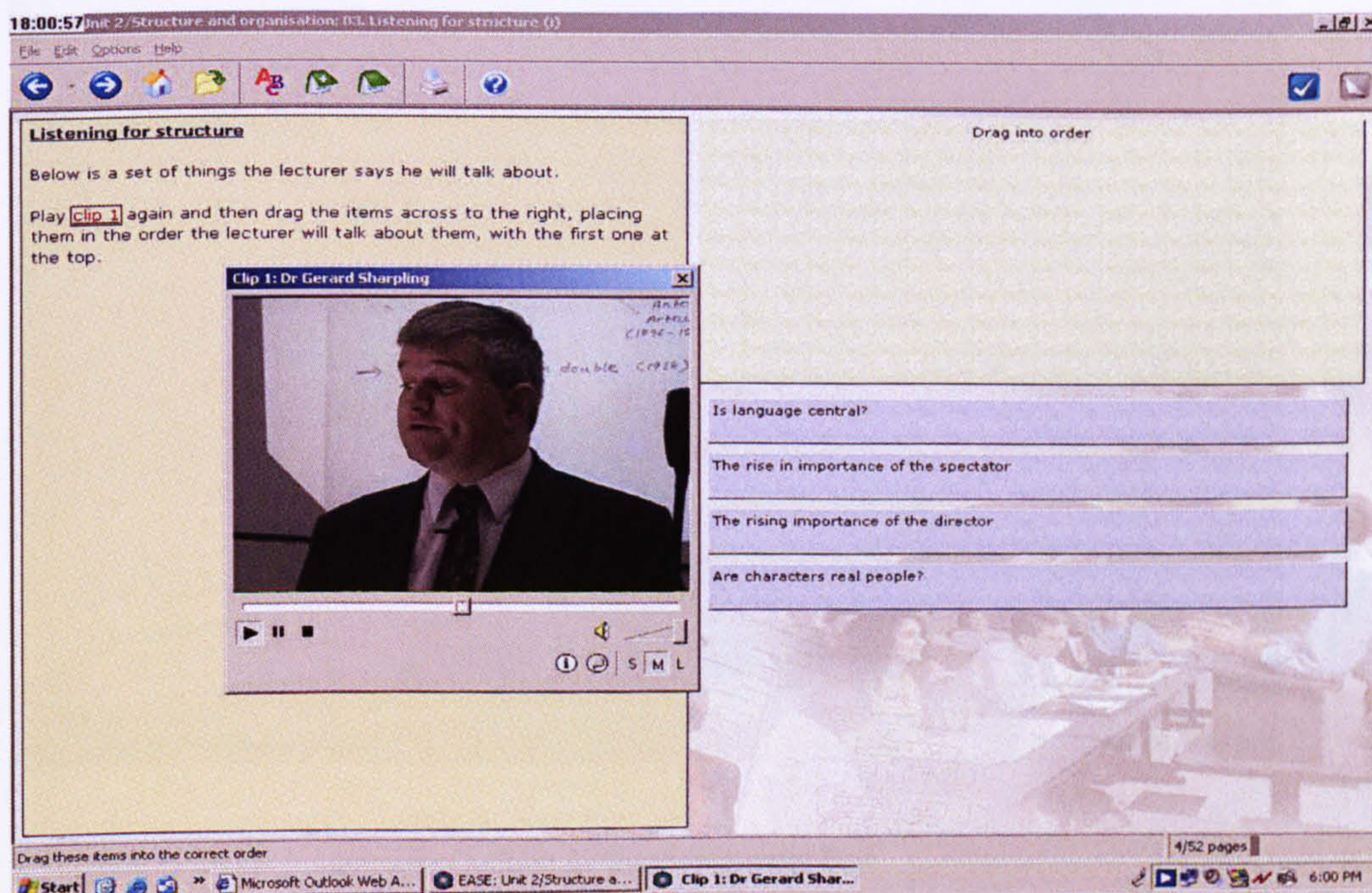
Screenshot 1.1: EASE *Listening to Lectures* Unit 1



Screenshot 1.2 : A page from *Listening to Lectures* Unit 1



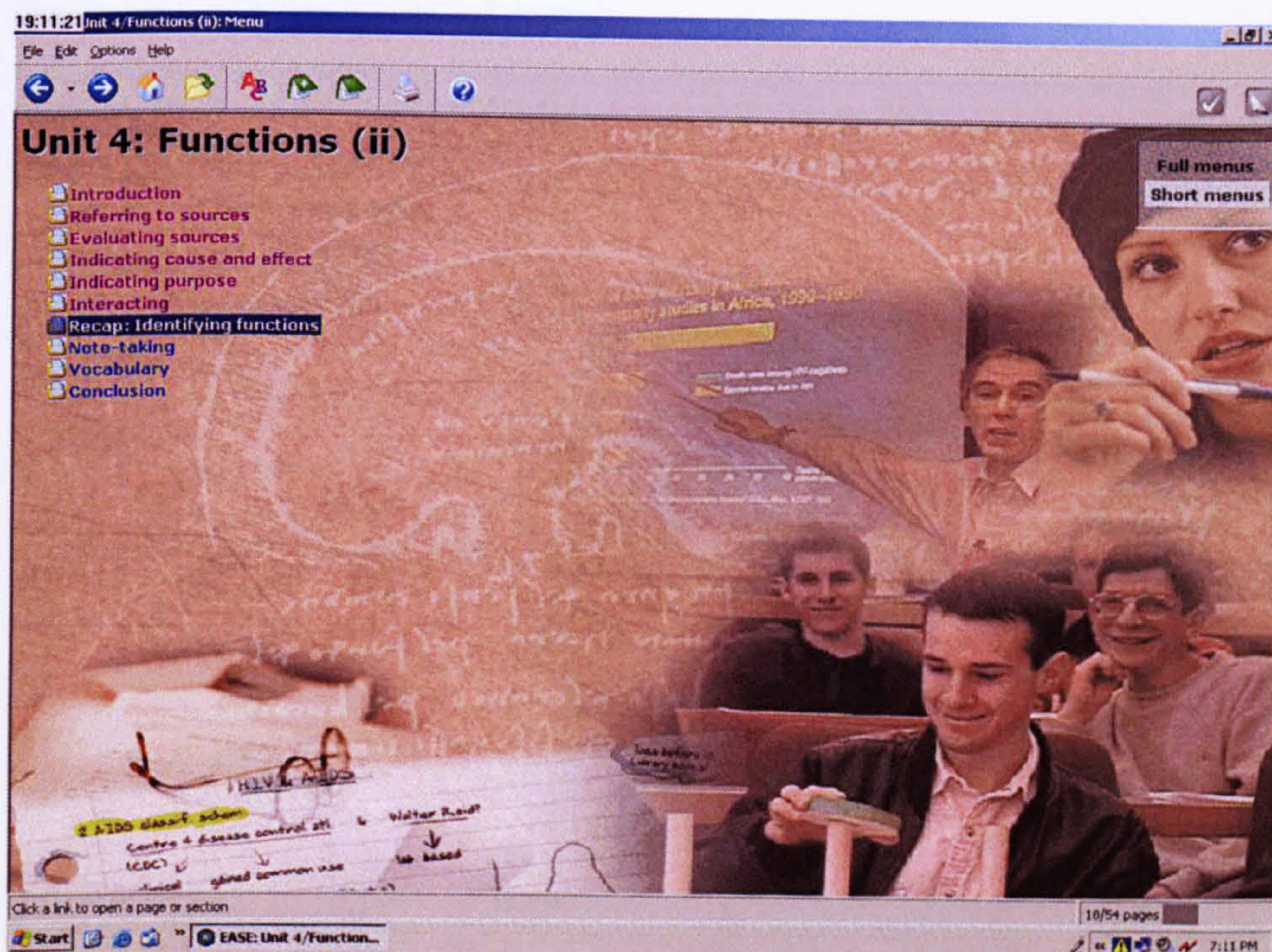
Screenshot 1.3: Unit 2 opening page



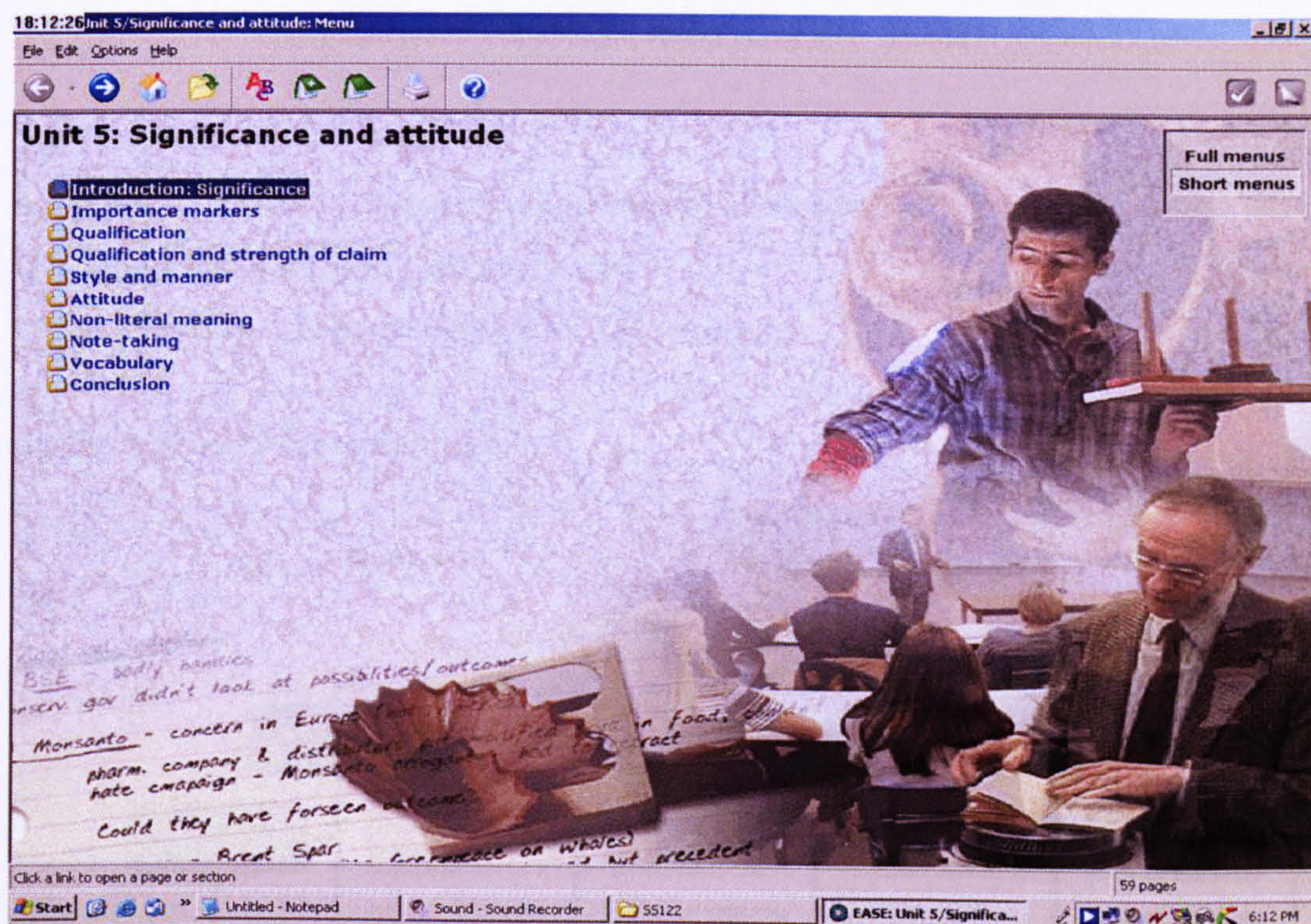
Screenshot 1.4: A page from unit 2



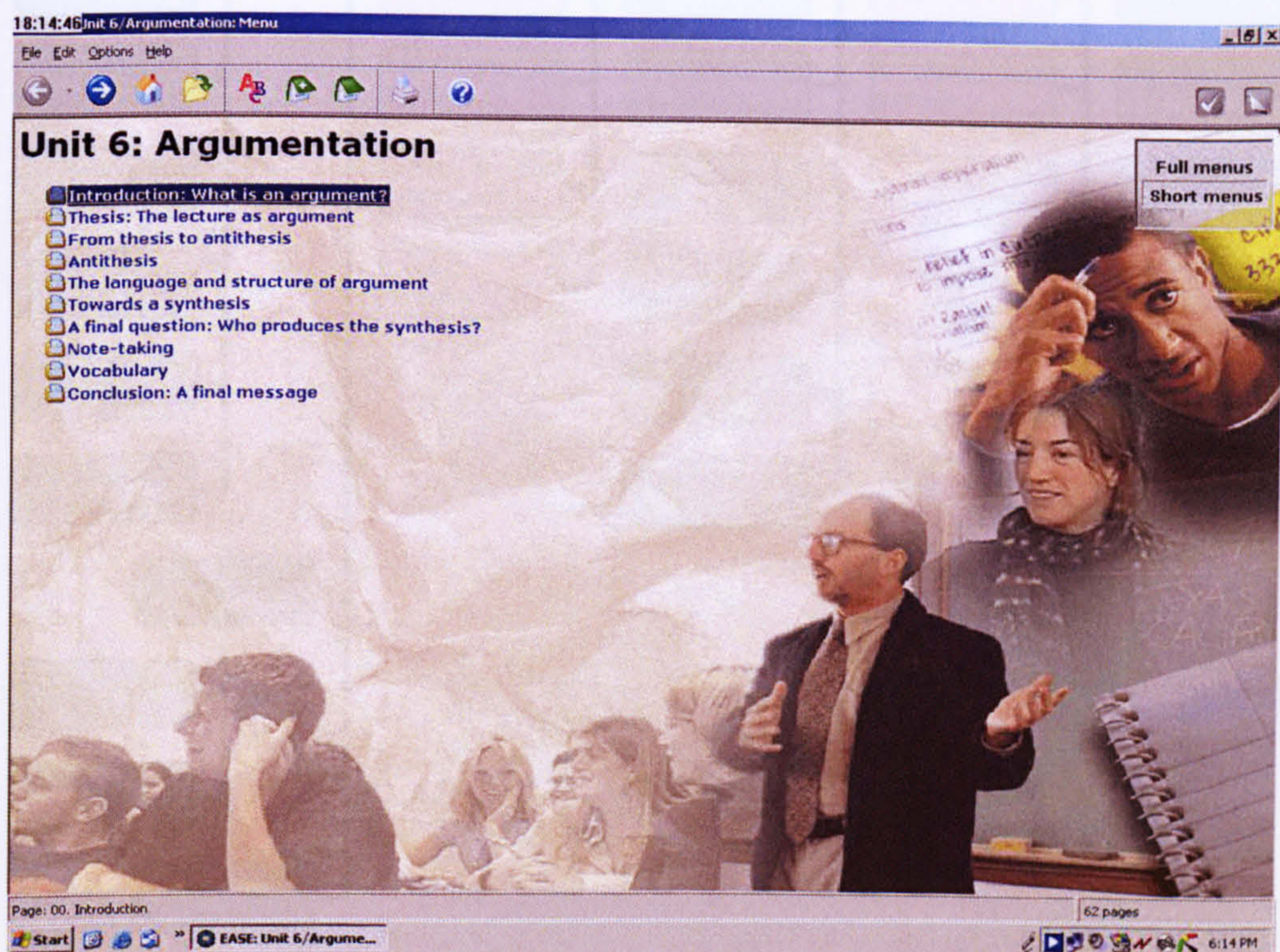
Screenshot 1.5 Unit 3 Functions Opening Page



Screenshot 1.6 Unit 4 Functions Opening Page



Screenshot Unit 5 Significance and Attitude



Screenshot 1.8 Unit 6 Argumentation of Listening to Lectures

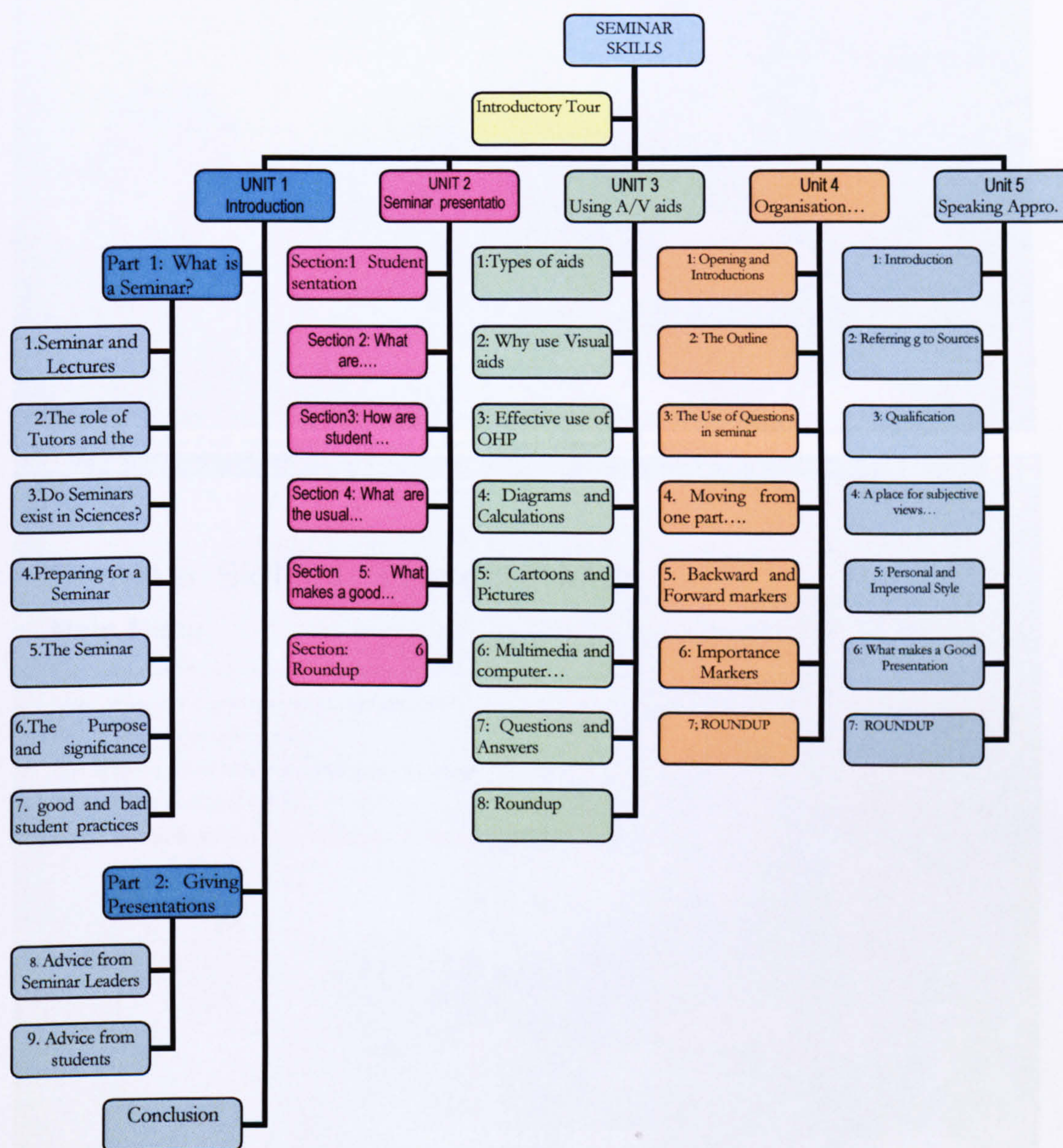
APPENDIX 1.2 DESCRIPTION OF SEMINAR SKILLS

Appendix 1.2: Description and Appearance of Materials

Description of Materials: *Seminar Skills*

Organisation chart of EASE Seminar Skills1: Presentations

See Pictures below for more detail.



Opening:

17:56:41 Seminar Skills: Main Menu

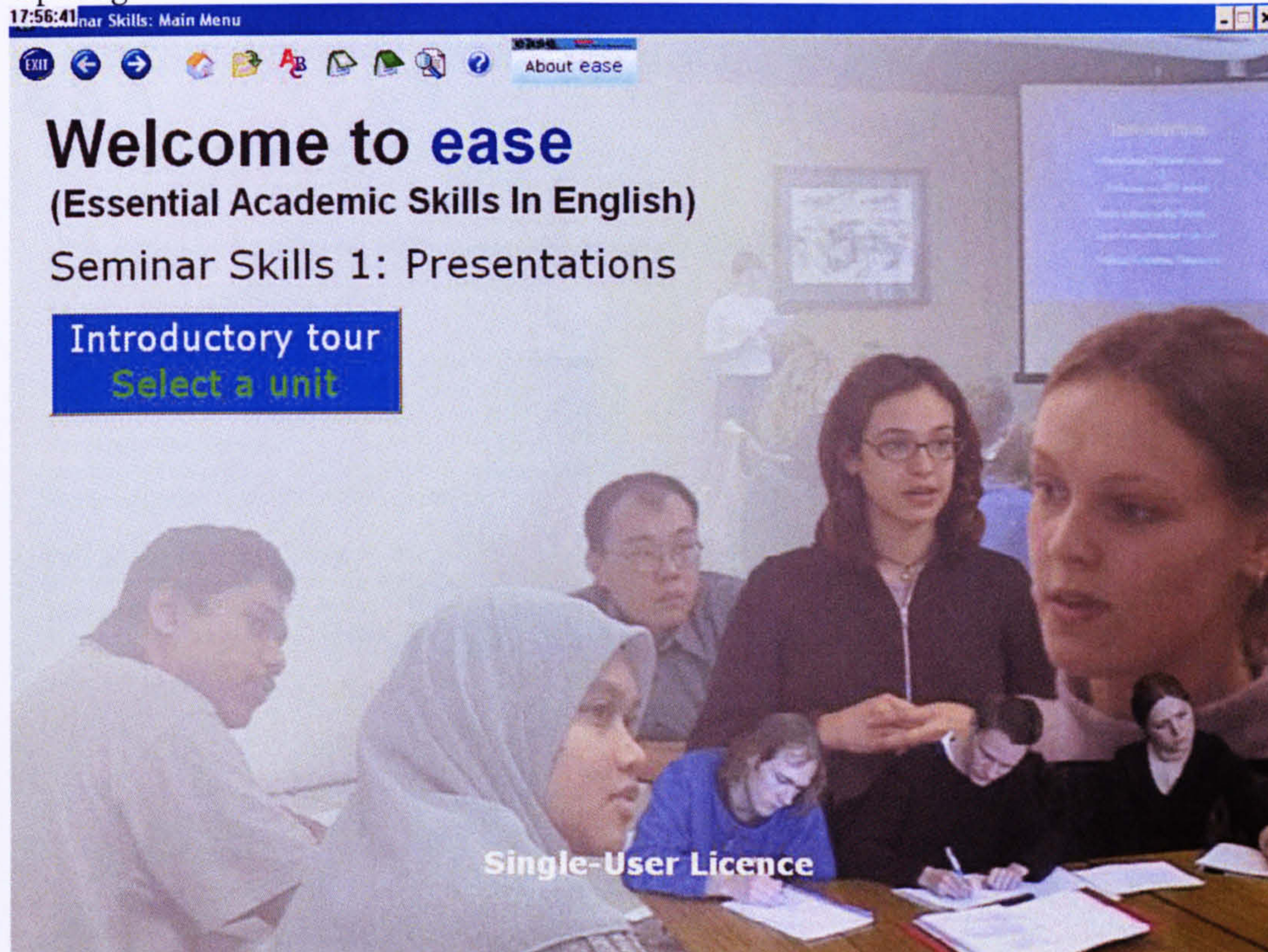
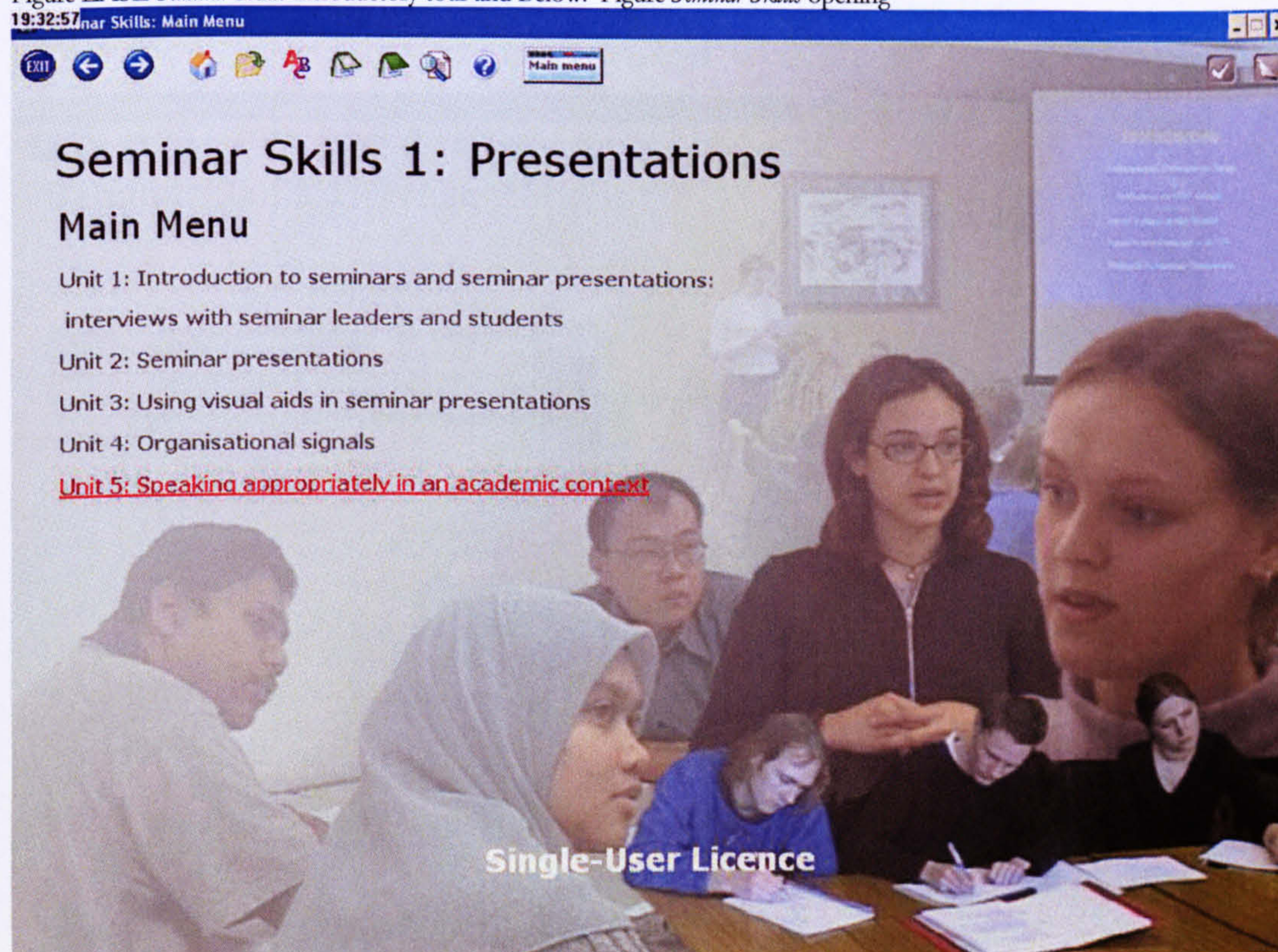


Figure *EASE Seminar Skills* introductory tour and Below: Figure *Seminar Skills* opening

19:32:57 Seminar Skills: Main Menu



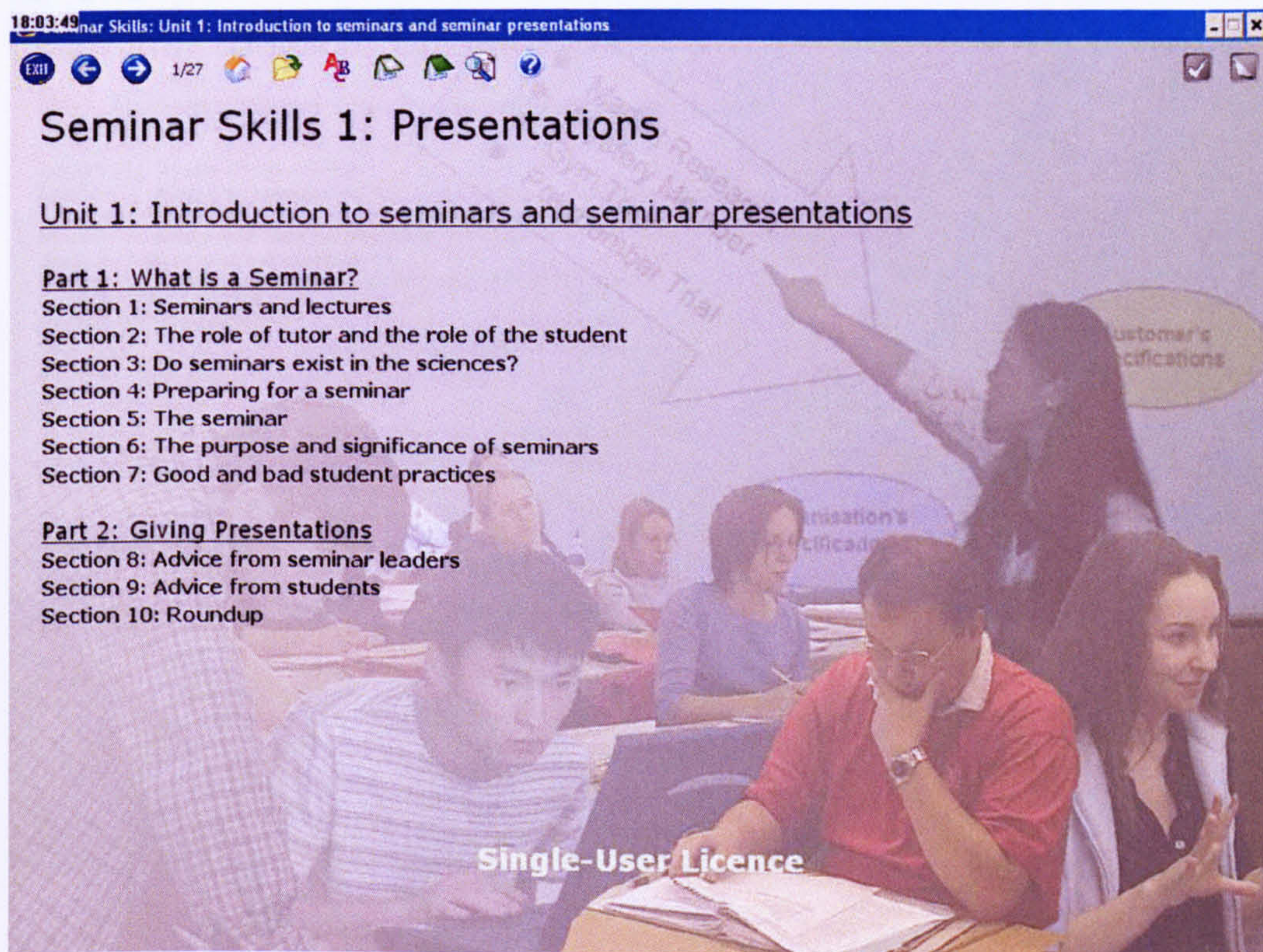
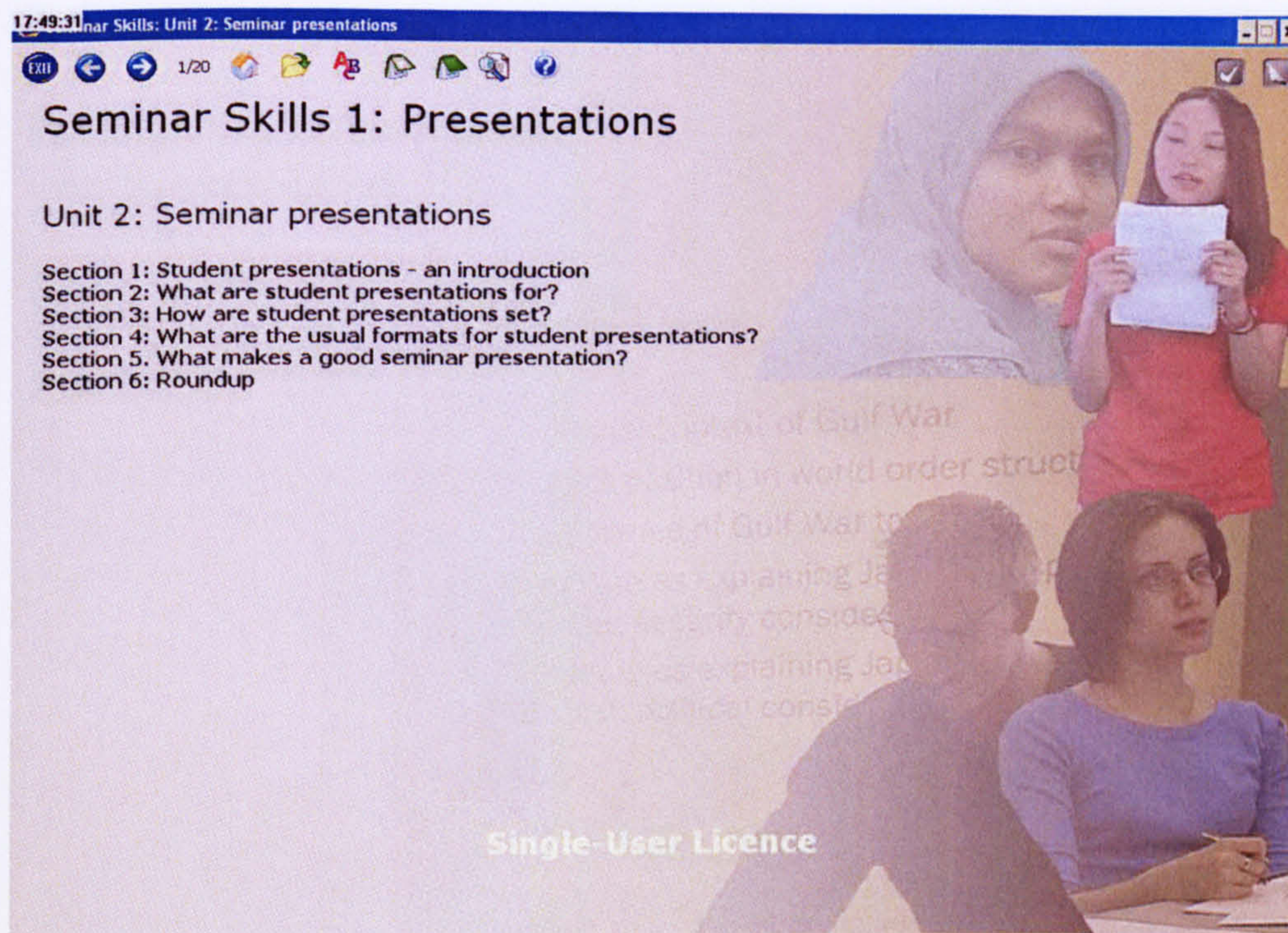


Figure EASE Seminar Skills Unit 1 and below Figure :EASE Seminar Skills Unit 2



18:59:11 Seminar Skills: Unit 3: Using visual aids in seminar presentations

EXIT 1/61

Seminar Skills 1: Presentations

Unit 3: Using visual aids in seminar presentations

- Section 1: Types of visual aids
- Section 2: Why use visual aids?
- Section 3: Effective use of the overhead projector
- Section 4: Diagrams and calculations
- Section 5: Cartoons and pictures
- Section 6: Multimedia and computer-aided presentations
- Section 7: Questions and answers
- Section 8: Roundup

Single-User Licence

18:19:28 Seminar Skills: Unit 4: Organisational Signals

EXIT 1/36

Seminar Skills 1: Presentations

Unit 4: Organisational Signals

Section 1: Openings and introductions

- Section 2: The outline
- Section 3: The use of questions in seminar presentations
- Section 4: Moving from one part of your presentation to another
- Section 5: Backward and forward markers
- Section 6: Importance markers
- Section 7: Roundup

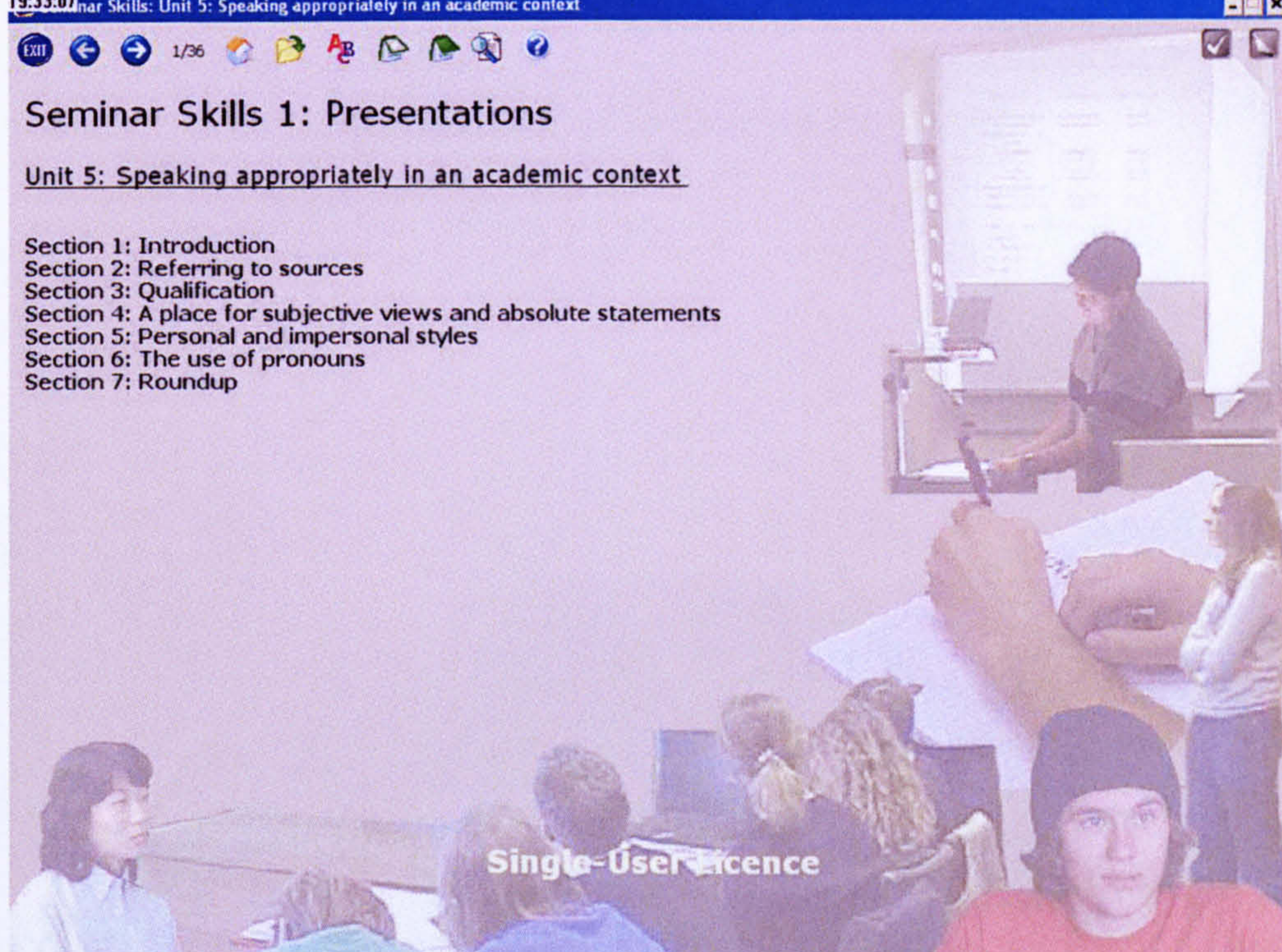
Single-User Licence



Seminar Skills 1: Presentations

Unit 5: Speaking appropriately in an academic context

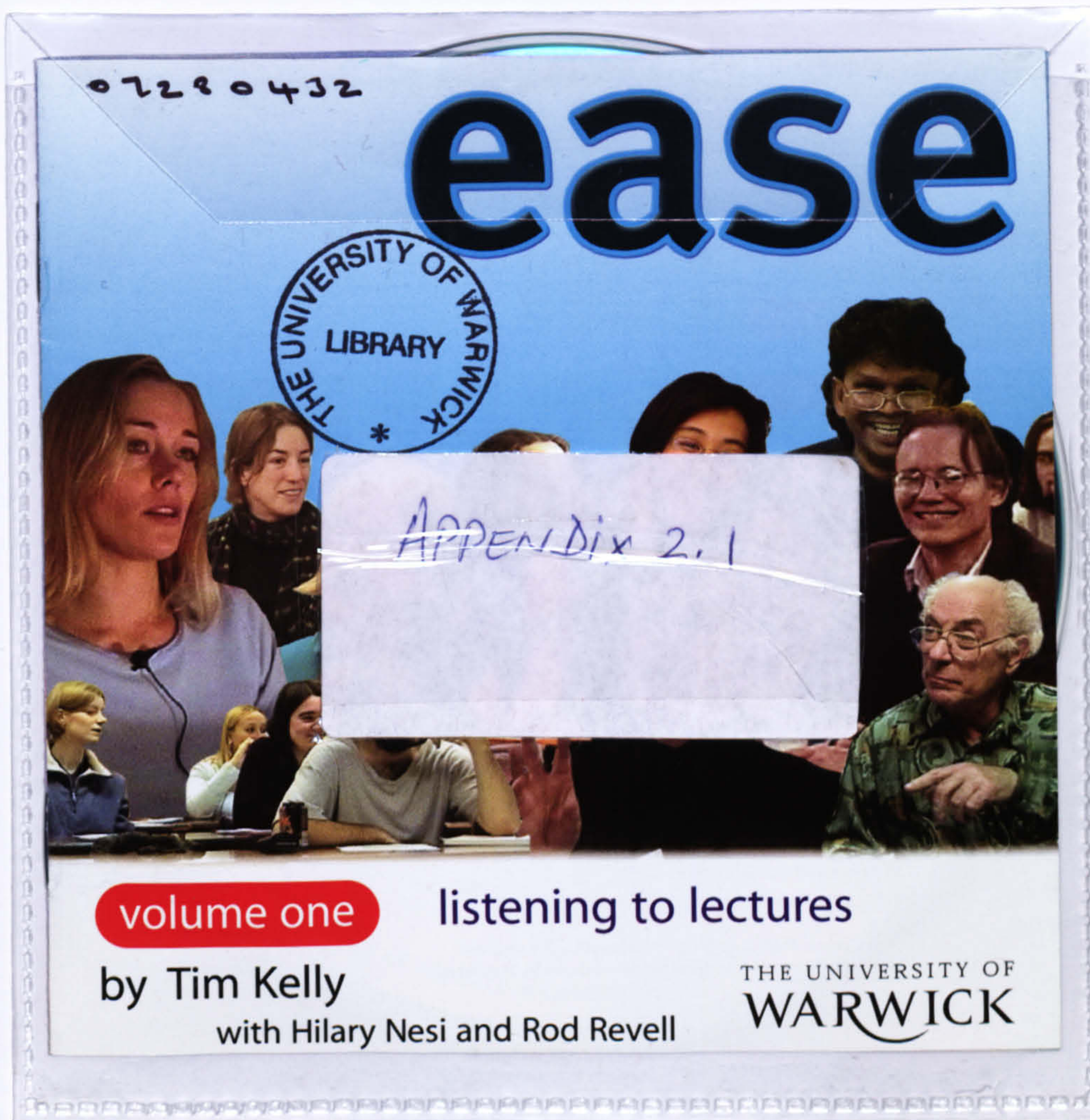
- Section 1: Introduction
- Section 2: Referring to sources
- Section 3: Qualification
- Section 4: A place for subjective views and absolute statements
- Section 5: Personal and impersonal styles
- Section 6: The use of pronouns
- Section 7: Roundup



APPENDIX 2

APPENDIX 2.1 EASE LISTENING TO LECTURES CD-ROM

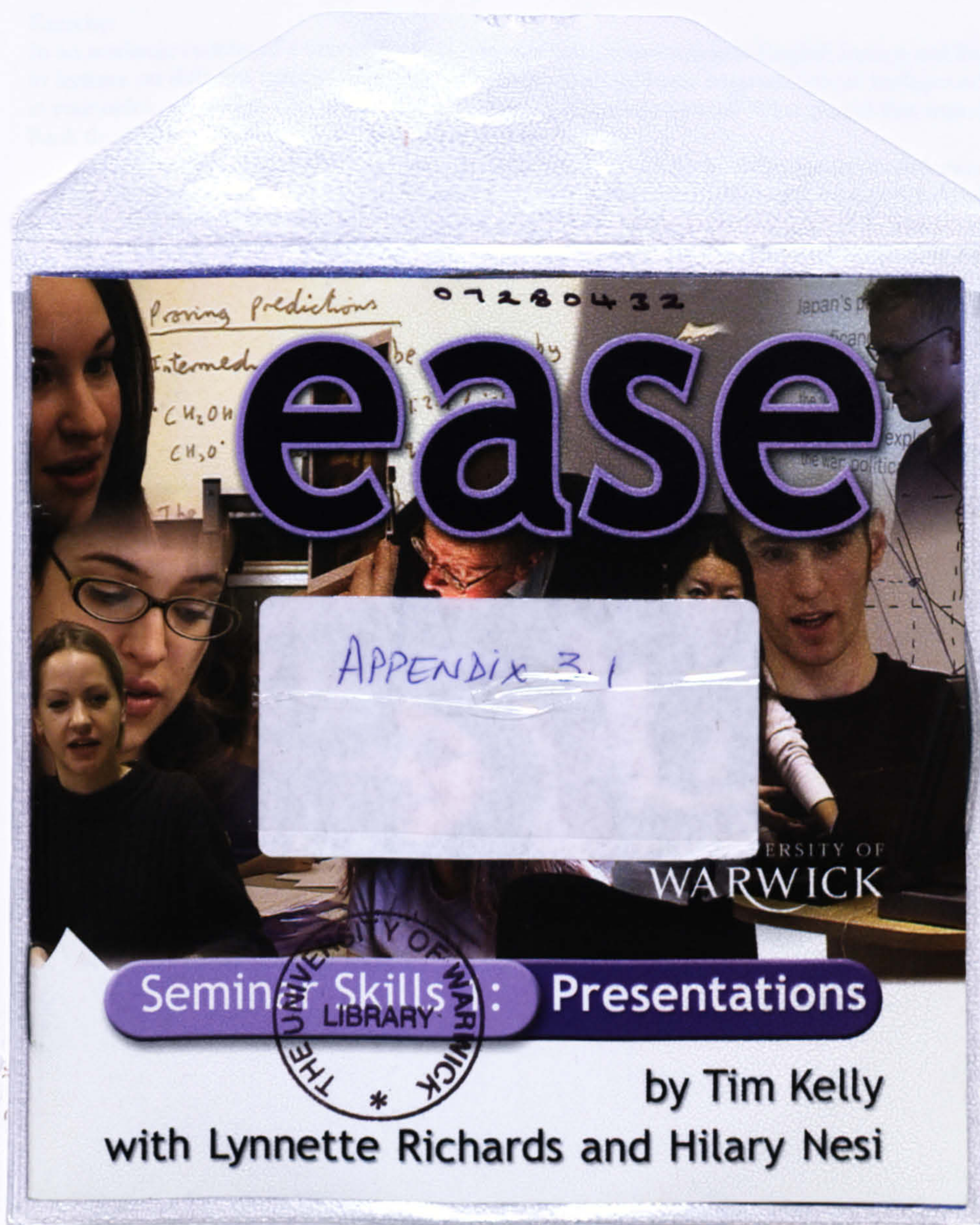
See Accompanying CD-ROM



APPENDIX 3

APPENDIX 3.1 EASE SEMINAR SKILLS CD-ROM

See Accompanying CD-ROM



APPENDIX 4

APPENDIX 4.1 FOCUSING EXERCISE

Focusing Exercise for Working with Multimedia Language Learning Materials to improve academic listening and speaking. Statements developed by the Researcher (Saima N. Sherazi).

Exercise:

In an academic context of a university where the medium of instruction is English foreign and local students have to listen to lectures on different subjects delivered by lecturers from different linguistic/ethnic backgrounds and nationalities. What in your opinion a student should look for in preparatory learning material? What should they include?

Rank the following statements in order of importance.

- A. Materials that provide an opportunity to familiarise your ear to different accents and ways of speaking.
 - B. Multicultural classes and faculty are the norm in most universities and a student should adapt to this. Materials that take this into account.
 - C. Materials that give adequate practice in academic skills particularly the skill of listening and speaking.
 - D. Materials that give adequate practice in grammar, reading and writing.
 - E. Materials that develop cognitive abilities by setting problems and tasks for students to develop the skill of listening because most university teaching is lecture based.
 - F. Materials that teach note taking, paraphrasing and summarising skills.
 - G. Materials that involve the whole person by including stories and humorous anecdotes that lecturers would use in their lectures.
 - H. Materials that provide immediate feedback on learner's errors and suggest additional materials through hot links embedded in the CD.
-

APPENDIX 4.2 POINTS FROM FG DISCUSSION

Points made by participants from which themes and categories were drawn.

Improvement of listening skills.

The order of what is important.

We are in a more advantaged position because we have studied English from the beginning.

Relating to humour. Get jokes Intrinsic to human nature.

Familiarize students with the accent.

Student interaction vs lecturer interaction. Student perspective vs Lecturer perspective

Too easy for this level didn't develop skills slight polishing of the skills we had.

English is almost like a first language.

Over emphasised the grammar reading and writing.

Ease would be brilliant for someone

Subconscious things were brought out we were not aware of.

We assume we know, lecture organisation of because we were forced to look at them. We became aware.

It was worthwhile.

While you are listening ou

Not look at it from students perspective

Primary point peripheral point.

Student closer to the lecturer's point of view same wavelength would be best of.

Signposting tonal variation we need to know we need to improve

Look at it from the student's perspective.

What CD is trying to

Concrete tips how to improve listening should be given.

Logical structure is fine. User friendly interface.

Sound effects with the exercise.

Monotonous slow periods in the CD

Broke your rhythm and momentum.

Review exercises review questions.

Love those vocabulary exercises
Vocabulary exercises were a recap for some not for us.
Big flaw: space, full stop, added to a right answer, the program wouldn't accept there should be something to say the spelling is wrong.
There is a Logical error in the programming.
If you don't type in the answer and press the DONE button thrice you get right answer
Button on the screen that says show the right answer.
As far as the sequence goes, like for example the note taking exercise I did it at the end.
2nd last exercise was note taking.
For assistance in vocabulary there was the dictionary similarly for grammar in the listening exercise... (whoever wants remedial help with grammar can consult this section)
Nice and fun it was vocabulary
The momentum we were all pumped up.(On a high and felt a come down with the level of the exercises.)
As far as the sequential thing is. Note taking should be moved up.
So if we do the note taking in the middle we would be able to manage time more effectively and efficiently time wise.
The lecture clip was so short that we could almost do the exercise because such a short time had passed we could do them without consulting our notes.
I am convinced you haven't done unit 5 and 6 because they are so comprehensive they gave us background information on the lectures.
Note taking exercises are boring and t least 13 minutes long they would least expect us to be entertained by the lecturer.
Didn't test the note taking ability.
You could view listen to the lecture one day and do the exercises the next day.
We did have incorrect answers but they were not challenging.
Most of us have done TOEFL and SAT scores this CD wasn't HARD for us.
The point is it wasn't challenging. It gets boring.
Seminar skills all new for us. Giving us real examples.
Much more relevant to our needs.
Freshman initiated into seminar skills.
Paraphrasing and summarising done in school.
One of the things referencing quoting we are exposed to at University.
Structural format CD does not give you a choice of levels.
SAT diagnostic test what we need help in.
Medium level Difficulty level should keep rising
No drastic change from unit to unit.
Learning (rationalizing the design of the CD)
Was this designed for British Students?
Environment of the CD?
I actually fell asleep.
It should have been colorful.
There should have been some sound effects.
Seminar Skills was much better.
More visual aids.
Resolution on seminar skills was better, the colors were better, sound was better.
Mr. Tim Kelly changed
The last message by him.
I actually liked the visual content of the first CD better yes it was a bit boring but we are at LUMS it's a fact of life that our lectures are boring. They are dry they are tedious and bit challenging to follow. Realistic selections. They were realistic selections.
What about the content of the lectures?
Topics under discussion were such that we could relate to them

APPENDIX 4.3 LONG TABLE

Focus Group 1. Long table Sample

Participants of Focus Group One (FG1) are anonymised as students A,B,C,D,E,F (f)andG and Focus Group Two (FG2) as H,J,K,L,M(f),N(f),O(f) with (f) denoting female gender.

79.	STUDENT D: it's like STUDENT G said English is almost like our first language so if you're considering a person who hasn't you know studied English much in the primary grades so the CD might help that person a bit more than us	English almost first language.mismatch		-	
80.	STUDENT E: I think they overemphasise the grammar and reading and writing they even had some exercises on the past tense perfect tense vocabulary for us that was [?]	Grammar overemphasised for these students		--	
81.	STUDENT G: for our needs it depends on the audience EASE would be brilliant for an audience which is not doesn't have the same perspectives with the learning experiences as us I mean it did help us but by bringing out the subconscious things to our attention and by doing that definitely I'd never thought of many many many things before it I did the program but then these things were I think they were already subconscious and you know	Awareness raising subconsciously done		+	
82.	SAIMA: What about I mean are you trying to say here you were already because I mean I've studied English myself from the very beginning and I felt that it made me aware of like for example how lecturers organise their lectures	Awareness of structure and Organisation			
83.	STUDENT B: okay basically I think there are a lot of things that we take for granted that we assume that we tend to know but despite this I think EASE to an extent obviously it brushed up our listening skills but I thought lecture organisation a lot of critical points that we tend to overlook were actually we were actually forced to actually look at that you were forced to see where the lecturer gave pauses where the lecturer was laying stress on where he was formal informative whatever different aspects of the lecture were given importance so that essence I think goes on to say a lot about EASE because EASE we might think it's very easy but I think on a subconscious level we've improved a lot and there are a lot of things that we just took for granted that are just coming to light now	Lecture organisation awareness		+	
84.	UNKNOWN: It was it was worthwhile nobody's saying it's not	Organisation important		+	
85.	STUDENT D: Don't you think that STUDENT B when you're writing down notes you're not supposed to notice that what the lecturer's actually doing you know when he's actually	Disagreement with the above		-	
86.	STUDENT F(F): you're not even supposed to you're not thinking on that level you're more interested what you're writing down and you are looking at signal phrases and how he's [?]	Organisation awareness not important		-	
87.	STUDENT D: exactly what he's saying rather than how he's saying				
88.	STUDENT C: I don't think you're supposed to look at it from the lecturer's side because you're not giving lectures for then next ten twenty years actually we are looking from the student's perspective how do we gain from that lecture	Lecture organisation not important to students		-	
89.	STUDENT B: okay even if you do look at it from a student's perspective you've got to know where the lecturer is laying stress on or which point is an important one which has peripheral importance and which is of primary importance okay then with your note exercises you will obviously not improve to that level	Organisation Important for knowing what is important for the lecturer		+	
90.	(discussion, unclear)				
91.	STUDENT G: I think I can shed some more light on this there's an Eastern saying which says when somebody delivers a speech and a hundred men listen to it each man walks out with his own	Affirmation of Importance of Lecture		+	

	understanding so basically when you look at it from the lecturers point of view and you try and analyse it and break it down as you hear it you're probably going to walk out closer to what the lecturer's perspective is and that should and the who student walks out on the same page as the lecturer will be the best of like STUDENT B said at some point we might think a peripheral point was a main point but if we had noticed the pauses and the organisation the peripheral we might be giving emphasis to a point which a lecturer didn't think was of primary importance so thinking like the lecturer would help us be on the same page as him	organisation			
92	STUDENT B: That's okay apart from taking notes and everything we also at some point in time we have to give presentations we have to deliver not lectures we have to deliver not lectures we have to give speeches and everything okay and signposting pauses tonal variation everything comes from that and if you're not to know what is the correct procedure of doing it then obviously how do you expect us to improve and how do you expect us to implement incorporate all those factors that actually improve your speech to be a part of our speech	Organisation important for students to be aware of		+	
93	STUDENT A: But the point we're trying to make is that we should look at it with the students perspective and in real time you don't have time to analyse [?]	Disagreement with the above point. Not enough time in real life		-	
94	(discussion, unclear)				
95	STUDENT A: you just try to take notes and you just go on revising your own	As students you do your own thing		-	
96	STUDENT G: what EASE is trying to do is make it a reflex action you should [?]	EASE helping create awareness of structure		+	
97	(discussion, unclear)				
98	STUDENT C: I think the CD is basically to help you in listening to lectures not help you in preparing and giving lectures	Re focussing on the main objective		+	
99	STUDENT D: you don't have to break them down like this	Disagreement		-	
100	STUDENT C: the level of flexibility if you're teaching something as the EASE series supposed to teach us how do we improve listening our listening skills to lectures we're not going to learn that by listening to other people's lectures at [?] you're supposed to give us some concrete tips on how to improve our listening skills	Concrete tips to effective listening to lectures should have been given		-	
101	STUDENT A: and I guess if we even know ways of analysing where it's being analytical and informative we're not super humans that are going to analyse each and every word in real time [?]	Unrealistic expectations raised		-	
102	STUDENT D: but if you're paying more attention to what the lecturer is actually saying rather than how he's saying	Focus on content not on style		+/-	
103	(discussion, unclear)				
104	SAIMA: okay STUDENT A let's just sort of move on from there we were talking about having a clear whether the CD had a clear focus now do you think the instructions are explicit the instructions for exercises for	Clear focus /Explicit Instruction			
105	ALL: yes				
106	STUDENT F(F): simple language easy to understand	Simple language		+	
107	SAIMA: what about the way the materials are organised and the sequence				
108	STUDENT F(F): the logical structure is fine	Logical structure		+	
109	STUDENT B: the transition is very smooth from one section to another that's not a problem	Smooth transition		+	

11	STUDENT F(F): you don't see that there's a big like a jump in the middle in the	Smooth transition		+	
11	UNKNOWN: A very user-friendly interface	User friendly interface		+	
11	STUDENT B: But I thought at times that if there were a little bit of sound effects with the exercises it would have helped because they were really boring at times I'm sorry but it was really boring at times	No sound effects		-	
11	STUDENT F(F): but it becomes a bit monotonous the units can become monotonous because the [?] throughout the CD that become a bit monotonous so if there was a bit more interchanging in the middle that would be good	Monotonous. Variation in the middle		-	
11	STUDENT E: And there were a lot of slow periods like after the note-taking exercise that will generally be it you know you would complete all your major points but then you would have this really long vocabulary and grammar exercise and that just broke your rhythm that just broke the momentum basically bored you if instead they had some review exercises or review questions that would be much more useful than having those end of you know [?] and grammar	Slow periods broke rhythm. Suggestion for review exercises		-	
11	SAIMA: you mean the students didn't get an opportunity to sort of consolidate what they had learnt	No consolidation			
11	STUDENT E: yeah kind of			-	
11	STUDENT G: I loved those vocabulary exercises they were fun	Vocab exercises good		+	
11	UNKNOWN: you're such a sergeant major				
11	STUDENT G: and then the summary points covered it in the end	Vocab exercises summarised		+	
12	STUDENT F(F): no the basic thing is that I'm sure we all found that we already know everything in the vocabulary exercises so that's why we found it monotonous someone who doesn't know it for them it probably was a recap for everything that went on in the unit it just wasn't for us	Vocab Monotonous		-	
12	SAIMA: but don't you I mean didn't you weren't you sort of surprised at times that there were some vocabulary items that you'd sort of				
12	STUDENT D: some of them they belong to English slang if you ask me	Slang expressions not picked up			
12	SAIMA: they belong to the English slang those are the ones that you didn't pick up				
12	STUDENT D: Yes				
12	SAIMA: but that's all the more reason you know you have to if you want to go to England and study in England this would sort of prepare you wouldn't it				
12	STUDENT B: at times the vocabulary exercises were annoying as well because there were some sophisticated terminologies used in the lecture which obviously at least I didn't know the spellings of them so whenever I typed them it was just incorrect and it would be an E or an A missing that's a big flaw and even if you put a full stop at the end it puts that it's incorrect	Vocab exercises annoying because domain specific vocab the system did not accept slight variation in spellings		-	
12	STUDENT C: yes in fact a space I mean instead of a single space being two words for example there if you mistakenly press two spaces that's something wrong with the programming of it actually	Agreement with the above. Programmed only to accept the exact answer and not		-	

		variation			
128	STUDENT A: yeah I would call that a logical error in the programming of the EASE	Logical error in programming		–	
129	STUDENT G: when it requires you to type in an answer you may type in the right answer but you might put a full stop at the end or a space before or you might be missing a vowel and your answer will be	Full stops or spaces counted as errors programming error		–	
130	STUDENT F(F): rejected				
131	STUDENT G: yeah exactly so it's annoying that's all it's just	Annoying		–	
132	STUDENT F(F): because in your brain you've given the right answer and you can't really find and you're looking at it and you're like what did I do wrong so there should be some way to pinpoint that okay you know the spelling's wrong or something like that	Agreeing with the above point. Some way of knowing that the answer is correct but the spelling is wrong		–	
133	STUDENT C: by the way I figured this out a little bit later if you don't type in an answer for example if press the done button twice it will give you the answer just like that I mean sometimes I didn't realise it was wrong so I just pressed the done button twice and it gave me the answer	Trick was to press done button twice for the right answer		–	

APPENDIX 4.4 FOCUS GROUP COLOUR CODED SAMPLE

Focus group 1 and 2 on EASE

Lahore University of Management Sciences – Pakistan

Saima Sherazi

Themes	Categories	Definition
Learning Effectiveness Richness	Appropriateness	Relevance of the CD to the needs of students
	Content	Imbalance in level, and treatment or sequence of certain instructional content.
	Motivation	Self evaluation mechanisms like scores to assess progress, authentic content
Usability Concerns	Competence	Functional features of the CD navigation, online help, adaptability.
	Suggestions	Suggestions for improving the program
	Motivation	Features in design that affect motivation
Affective Dimensions	Management	Managing the program in the context of the course
	Justification	Rationalisation of negative criticism
	Motivation	Motivational and affective dimensions of authentic cultural exposure

75.	SAIMA: alright can you mark it on your piece of paper then we can start with now if you were to apply all this to the CD's that you have worked on right which is the EASE CD and do you think yeah sorry do you think the aims and objectives of the first CD let's talk about the first CD first which is listening to lectures they have been achieved do you think the CD has a clear focus
76.	STUDENT A: that was actually too easy for this level for being taught at this level that didn't really develop skills like as in what I mean to say is that like yeah it was a slight polishing of the abilities we already had so that was not like a [?] program that we went to
77.	SAIMA: but STUDENT A were you had you been exposed to something like that was your listening ever tested mean
78.	STUDENT A: no it was not tested before but
79.	STUDENT D: it's like STUDENT G said English is almost like our first language so if you're considering a person who hasn't you know studied English much in the primary grades so the CD might help that person a bit more than us
80.	STUDENT E: I think they overemphasise the grammar and reading and writing they even had some exercises on the past tense perfect tense vocabulary for us that was [?]

81.	STUDENT G: for our needs it depends on the audience EASE would be brilliant for an audience which is not doesn't have the same perspectives with the learning experiences as us I mean it did help us but by bringing out the subconscious things to our attention and by doing that definitely I'd never thought of many many many things before it I did the program but then these things were I think they were already subconscious and you know
82.	SAIMA: What about I mean are you trying to say here you were already because I mean I've studied English myself from the very beginning and I felt that it made me aware of like for example how lecturers organise their lectures
83.	STUDENT B: okay basically I think there are a lot of things that we take for granted that we assume that we tend to know but despite this I think EASE to an extent obviously it brushed up our listening skills but I thought lecture organisation a lot of critical points that we tend to overlook were actually we were actually forced to actually look at that you were forced to see where the lecturer gave pauses where the lecturer was laying stress on where he was formal informative whatever different aspects of the lecture were given importance so that essence I think goes on to say a lot about EASE because EASE we might think it's very easy but I think on a subconscious level we've improved a lot and there are a lot of things that we just took for granted that are just coming to light now
84.	UNKNOWN: It was it was worthwhile nobody's saying it's not
85.	STUDENT D: Don't you think that STUDENT B when you're writing down notes you're not supposed to notice that what the lecturer's actually doing you know when he's actually
86.	STUDENT F(F): you're not even supposed to you're not thinking on that level you're more interested what you're writing down and you are looking at signal phrases and how he's [?]
87.	STUDENT D: exactly what he's saying rather than how he's saying
88.	STUDENT C: I don't think you're supposed to look at it from the lecturer's side because you're not giving lectures for then next ten twenty years actually we are looking from the student's perspective how do we gain from that lecture
89.	STUDENT B: okay even if you do look at it from a student's perspective you've got to know where the lecturer is laying stress on or which point is an important one which has peripheral importance and which is of primary importance okay then with your note exercises you will obviously not improve to that level
90.	(discussion, unclear)
91.	STUDENT G: I think I can shed some more light on this there's an Eastern saying which says when somebody delivers a speech and a hundred men listen to it each man walks out with his own understanding so basically when you look at it from the lecturers point of view and you try and analyse it and break it down as you hear it you're probably going to walk out closer to what the lecturer's perspective is and that should and the who student walks out on the same page as the lecturer will be the best of like STUDENT B said at some point we might think a peripheral point was a main point but if we had noticed the pauses and the organisation the peripheral we might be giving emphasis to a point which a lecturer didn't think was of primary importance so thinking like the lecturer would help us be on the same page as him
92.	STUDENT B: That's okay apart from taking notes and everything we also at some point in time we have to give presentations we have to deliver not lectures we have to deliver not lectures we have to give speeches and everything okay and signposting pauses tonal variation everything comes from that and if you're not to know what is the correct procedure of doing it then obviously how do you expect us to improve and how do you expect us to implement incorporate all those factors that actually improve your speech to be a part of our speech
93.	STUDENT A: But the point we're trying to make is that we should look at it with the students perspective and in real time you don't have time to analyse [?]
94.	(discussion, unclear)
95.	STUDENT A: you just try to take notes and you just go on revising your own
96.	STUDENT G: what EASE is trying to do is make it a reflex action you should [?]
97.	(discussion, unclear)
98.	STUDENT C: I think the CD is basically to help you in listening to lectures not help you in preparing and giving lectures
99.	STUDENT D: you don't have to break them down like this
100.	STUDENT C: the level of flexibility if you're teaching something as the EASE series supposed to teach us how do we improve listening our listening skills to lectures we're not going to learn that by listening to other people's lectures at [?] you're supposed to give us some concrete tips on how to improve our listening skills

101.	STUDENT A: and I guess if we even know ways of analysing where it's being analytical and informative we're not super humans that are going to analyse each and every word in real time [?]
102.	STUDENT D: but if you're paying more attention to what the lecturer is actually saying rather than how he's saying
103.	(discussion, unclear)
104.	SAIMA: okay STUDENT A let's just sort of move on from there we were talking about having a clear whether the CD had a clear focus now do you think the instructions are explicit the instructions for exercises for
105.	ALL: yes
106.	STUDENT F(F): simple language easy to understand
107.	SAIMA: what about the way the materials are organised and the sequence
108.	STUDENT F(F): the logical structure is fine
109.	STUDENT B: the transition is very smooth from one section to another that's not a problem
110.	STUDENT F(F): you don't see that there's a big like a jump in the middle in the
111.	UNKNOWN: A very user-friendly interface
112.	STUDENT B: But I thought at times that if there were a little bit of sound effects with the exercises it would have helped because they were really boring at times I'm sorry but it was really boring at times
113.	STUDENT F(F): but it becomes a bit monotonous the units can become monotonous because the [?] throughout the CD that become a bit monotonous so if there was a bit more interchanging in the middle that would be good
114.	STUDENT E: And there were a lot of slow periods like after the note-taking exercise that will generally be it you know you would complete all your major points but then you would have this really long vocabulary and grammar exercise and that just broke your rhythm that just broke the momentum basically bored you if instead they had some review exercises or review questions that would be much more useful than having those end of you know [?] and grammar
115.	SAIMA: you mean the students didn't get an opportunity to sort of consolidate what they had learnt
116.	STUDENT E: yeah kind of
117.	STUDENT G: I loved those vocabulary exercises they were fun
118.	UNKNOWN: you're such a social science major
119.	STUDENT G: and then the summary points covered it in the end
120.	STUDENT F(F): no the basic thing is that I'm sure we all found that we already know everything in the vocabulary exercises so that's why we found it monotonous someone who doesn't know it for them it probably was a recap for everything that went on in the unit it just wasn't for us
121.	SAIMA: but don't you I mean didn't you weren't you sort of surprised at times that there were some vocabulary items that you'd sort of
122.	STUDENT D: some of them they belong to English slang if you ask me
123.	SAIMA: they belong to the English slang those are the ones that you didn't pick up
124.	STUDENT D: Yes
125.	SAIMA: but that's all the more reason you know you have to if you want to go to England and study in England this would sort of prepare you wouldn't it
126.	STUDENT B: at times the vocabulary exercises were annoying as well because there were some sophisticated terminologies used in the lecture which obviously at least I didn't know the spellings of them so whenever I typed them it was just incorrect and it would be an E or an A missing that's a big flaw and even if you put a full stop at the end it puts that it's incorrect
127.	STUDENT C: yes in fact a space I mean instead of a single space being two words for example there if you mistakenly press two spaces that's something wrong with the programming of it actually
128.	STUDENT A: yeah I would call that a logical error in the programming of the EASE
129.	STUDENT G: when it requires you to type in an answer you may type in the right answer but you might put a full stop at the end or a space before or you might be missing a vowel and your answer will be
130.	STUDENT F(F): rejected
131.	STUDENT G: yeah exactly so it's annoying that's all it's just

132.	STUDENT F(F): because in your brain you've given the right answer and you can't really find and you're looking at it and you're like what did I do wrong so there should be some way to pinpoint that okay you know the spelling's wrong or something like that
133.	STUDENT C: by the way I figured this out a little bit later if you don't type in an answer for example if press the done button twice it will give you the answer just like that I mean sometimes I didn't realise it was wrong so I just pressed the done button twice and it gave me the answer
134.	SAIMA: so you didn't go through the whole CD like that did you
135.	STUDENT C: no I did go through the CD properly
136.	STUDENT B: it actually said that that if you press it twice they'll give the correct answer automatically there were certain instructions but obviously [?]
137.	STUDENT F(F): and suppose there is a button on the screen that says show the right answer so if all else fails ? for next time
138.	STUDENT E: as far as the sequence goes like what I did the note-taking exercise would be in the middle but I'd always do it at the end but that's how I got
139.	STUDENT F(F): note-taking was second to last
140.	STUDENT E: second last I would do it at the end to review on my concepts because vocab was always separate
141.	SAIMA: did that help STUDENT E did that help in your getting the right answers
142.	STUDENT E: yeah I thought it helped me a lot and the vocab and the grammar thing I mean okay I think some students need vocab and grammar I accept that but it just felt like disconnected from the rest of the unit like when you did some listening skills and it's called grammar you should have a separate unit on grammar and vocabulary
143.	STUDENT F(F): so for whoever who wants to do that can do that part otherwise they can skip it
144.	STUDENT E: there was just no connection
145.	STUDENT G: like for assistance in vocabulary like in seminar skills you had a dictionary if in listening to lectures you had some sort of if you need that help if the need it
146.	SAIMA: if you need grammar work
147.	STUDENT G: if you don't need it then if you compulsorily force us to do it then it takes focus away from listening
148.	SAIMA: but you didn't feel that it was for the revision in the sense that it sort of made you aware of
149.	STUDENT G: not so much
150.	STUDENT F(F): i don't know how to describe it you were like on this peak after note-taking and it sort of liked slammed you down because it was all vocabulary
151.	SAIMA: didn't you feel it was giving you a respite
152.	STUDENT G: yes I did that's why I liked it so much because it was like nice and fun compared to taking you know note-taking
153.	(discussion, unclear)
154.	SAIMA: the whole idea was that here you are doing some very cerebral activity which is making you think and work very hard and then it is giving you a respite okay take a break breather
155.	STUDENT G: exactly that's what I thought
156.	STUDENT F(F): but after note-taking you are on this momentum that [?] so you're all pumped and you're like good good good
157.	STUDENT B: I think after the note-taking exercise okay basically through the sequential order I think there should be note-taking should be moved a little bit up because basically we see that clip a zillion times before we actually come to the note-taking thing so if we do the note-taking exercise in the middle of it then I think to an extent we'll be able to answer the rest of the exercises with just viewing the tape once and then obviously we'll be able to manage time more effectively and do it more comprehensively than we did I think
158.	STUDENT A: And there's another point that I'd like to raise I don't know whether they will agree with me or not like the note-taking the lecture from which we had to take notes was so short that our note-taking skills was not actually tested because the questions that followed we could just simply answer them it was such a short time had passed and we could just answer them without consulting our notes so we didn't exactly
159.	SAIMA: and yet sometimes you got them wrong
160.	(discussion, unclear)

161.	STUDENT A: but we didn't exactly get to know that we were making the right notes
162.	SAIMA: one at a time please
163.	STUDENT F(F): I'm certain that you have in unit 5 and 6 because their note-taking was very comprehensive and there was
164.	STUDENT B: it's actually ten to fifteen minutes long and it's actually a very comprehensive lecture which has tonnes of information
165.	STUDENT F(F): and it was so comprehensive
166.	SAIMA: you haven't done last two of
167.	STUDENT A: the second CD
168.	SAIMA: but have you done the first
169.	STUDENT A: yeah
170.	SAIMA: are you sure
171.	STUDENT F(F): unit 5 and 6 was so comprehensive that before the start of the lecture that they actually before the start of the lecture gave us background information on it so that just gives an idea of how comprehensive even they felt it was because in the first four units that so wasn't there
172.	SAIMA: so the form has changed slightly okay
173.	STUDENT B: But if all of us are complaining about being boring and being dry then if do you expect that the noting exercise is actually 30 minutes long would you be entertained by the lecturer
174.	STUDENT A: no I'm just trying to say that if it didn't test the note-taking ability exactly bit yeah

APPENDIX 4.5 NVIVO CODING REPORT OF FOCUS GROUP 2.

NVivo revision 2.0.161 Licensee: Saima Sherazi

Project: EASE User: Administrator Date: 1/29/2007 - 1:47:19 PM

DOCUMENT CODING REPORT

Document: Focus group on EASE 2-DH

Created: 12/22/2005 - 1:35:43 PM

Modified: 1/3/2006 - 3:53:52 PM

Description:

Focus group 2 on EASE

Node: (3 2) /Affective Dimensions/Justification

Passage 1 of 8 Section 0, Paras 429 to 437, 1099 chars.

429: SHEHERYAR SADIQ: alright look there's absolutely no pressure on doing that CD you don't have any pressure you don't have it's just a regular class right there's no pressure there's no exam coming up right so it's a kind of relief that alright yeah there was something funny in it so now it's got my attention but if I'd had an exam tomorrow right or in the next month it's a pretty long exam and obviously I don't have time

430:

431: DANIA NAVID KHAN: they're not teaching you humour that's a different [?]

432:

433: SHEHERYAR SADIQ: that's not what I meant it's basically aimed at keeping your attention about something and it helps but if it comes out unexpectedly then okay but if I have an exam coming up then I wouldn't go for something like that

434:

435: NADIR KHAN: it would be like that right now if you're having classes and you have an exam tomorrow and the tutor starts cracking a joke and it's like please get on with it

436:

437: HUMA MAHMOOD MOTIWALA: read it read it I think you've got the definition wrong it means humorous anecdotes which apply to the topic and you know including stories and anecdotes by making it

Passage 2 of 8
Section 0, Para 823, 146 chars.

823: HUMA MAHMOOD MOTIWALA: actually I think I learnt a lot of stuff that we actually take for granted this is what you realise or this is what happens

Passage 3 of 8
Section 0, Para 939, 320 chars.

939: HUMA MAHMOOD MOTIWALA: you don't really have time to think about these things and implement them but now that I'm thinking about it at that point I thought the fill in the blanks was really kind of futile but right now I look at my notes and there's a lot of words that I miss out so it really helps you to fill in those

Passage 4 of 8
Section 0, Paras 947 to 951, 601 chars.

947: SHEHERYAR SADIQ: firstly you fill the blanks regarding whether or not you actually understood the accent in which the lecturer was speaking probably was testing whether or not if you had that ability of you know finding out what he was saying and obviously it doesn't have the concern but this is basically focused on mainly students here regard everybody not everybody is like as good in English as

948:

949: USMAN ZAAFAR: yeah but [?] fill in the blanks aren't you testing a person's memory rather than his note-taking skill

950:

951: SHEHERYAR SADIQ: but they have the option of running through the clip again right

Passage 5 of 8
Section 0, Paras 1011 to 1015, 417 chars.

1011: USMAN ZAAFAR: maybe they should have asked us what you want to do maybe they should have said do you know this and do you know this and if you do then you can take the score and skip ahead

1012:

1013: HUMA MAHMOOD MOTIWALA: you don't always know what you know

1014:

1015: NADIR KHAN: you might think you know everything about everything something and in the end you'd find out that this is a good point here or what I'm doing here is wrong

Passage 6 of 8
Section 0, Para 1071, 48 chars.

1071: USMAN ZAAFAR: how difficult is "Ctrl+C" "Ctrl+V"

Passage 7 of 8
Section 0, Para 1149, 128 chars.

1149: NADIR KHAN: maybe we haven't realised that how much it has helped us and we're just you know like saying it's not helpful at all

Passage 8 of 8
Section 0, Paras 1243 to 1245, 237 chars.

1243: SHEHERYAR SADIQ: I said before that it robs us of the actual environment of the class

1244:

1245: DUA SHABBIR SYED: the shy students need that environment because they need to come out of that shell and this is just enclosing them in it even more

1246:

APPENDIX 4.6 SELECTED COMMENTS FROM FOCUS GROUPS

Appendix Chapter Four Selected Comments from Section 4.2.2.3

Appendix Table 4.5 Themes and Categories of Focus Groups

Themes	Categories	Definition
Learning Richness	Appropriateness	Relevance of the CD to the needs of students
	Content	Imbalance in level, and treatment or sequence of instructional content.
	Motivation	Self evaluation assessment mechanisms like scores to assess progress, and authentic content of video clips motivating
Improving Competence	Usability	Functional features of the CD navigation, online help, adaptability.
	Suggestions	Suggestions for improving the program
	Motivation	Features in design that affect motivation
Affective Dimensions	Management	Management of learning innovation
	Justification	Rationalisation of negative criticism
	Motivation	Motivational and affective dimensions of authentic cultural exposure

Appendix Table 4.6 Themes and Categories and Selected Student Comments

Themes	Categories	Definition
--------	------------	------------

Learning Richness	Appropriateness	<p>STUDENT D: it's like STUDENT G said English is almost like our first language so if you're considering a person who hasn't you know studied English much in the primary grades so the CD might help that person a bit more than us</p> <hr/> <p>STUDENT G: for our needs it depends on the audience EASE would be brilliant for an audience which is not doesn't have the same perspectives with the learning experiences as us I mean it did help us but by bringing out the subconscious things to our attention and by doing that definitely I'd never thought of many many many things before it I did the program but then these things were I think they were already subconscious and you know.(FG1)</p> <hr/> <p>STUDENT B: okay basically I think there are a lot of things that we take for granted that we assume that we tend to know but despite this I think EASE to an extent obviously it brushed up our listening skills but I thought lecture organisation a lot of critical points that we tend to overlook were actually we were actually forced to actually look at that you were forced to see where the lecturer gave pauses where the lecturer was laying stress on where he was formal informative whatever different aspects of the lecture were given importance so that essence I think goes on to say a lot about EASE because EASE we might think it's very easy but I think on a subconscious level we've improved a lot and there are a lot of things that we just took for granted that are just coming to light now</p> <hr/> <p>STUDENT G: you know sprinkled all kinds of local flavour in it and English English is quaint and funny and quirky and everybody knows that and once you've mastered that idiom then you know you're qualified to go and study there and that's one of the advantages that it gives you an insight into the quiriness</p> <hr/> <p>STUDENT B: and I think one thing for sure that all of us hopefully after our Bachelor's at some point in time we will be hopefully going abroad and probably some of us for studies for studying as in studying abroad so I think at that point in time this thing will obviously help us because then although I'm sure by that time three years later all of us will have forgotten about this but still as in if we were to go right after this CD abroad I think it would help massively because then we would know what dialects and what accents to expect at least in a British university so I think that's one of the unexpected things that have improved</p> <hr/>
-------------------	-----------------	---

	Content	<p>STUDENT F (F): but after note-taking you are on this momentum that [?] so you're all pumped and you're like good good good</p> <hr/> <p>STUDENT O (F): what really kills you on CD one is that okay you're all tensed-up you know lecture lecture you sit listening carefully and like after two minutes of something world economic theory the question is how would you spell the world</p> <hr/> <p>STUDENT N (F):the exercises didn't relate to the lecture</p> <hr/> <p><i>STUDENT L: [?] I insisted again and again that the level of the CD isn't really getting beyond a certain point it's pretty much staying at the same stagnant level except for I think the last unit.</i></p> <hr/> <p><i>STUDENT L: no but the thing is it teaches you exactly what content is necessary you know when you talk about listening to lectures like I mean the first two units I have done on CD two already it says you know find the appropriate material on the site so basically CD one teaches you how to grasp the essence of the lectures exactly what the person is saying so yeah to an extent it is important as far as improvement is concerned well to an extent yes I guess perhaps I mean unit six teaches antithesis and that taught us something but on the whole not really the level of progress I would have expected</i></p> <hr/> <p>STUDENT H: the first CD it was basically skills we had already learnt in the A-levels and O-levels and all so I mean we didn't find it as challenging or we didn't learn as much as we could CD two on the other hand I mean basically no-one's really taught us how to do a presentation properly so I mean [?] you're like oh this is what you do you know this is how you do it very detailed information</p> <hr/> <p>STUDENT F (F): but it becomes a bit monotonous the units can become monotonous because the [?] throughout the CD that become a bit monotonous so if there was a bit more interchanging in the middle that would be good</p> <hr/> <p>STUDENT E: And there were a lot of slow periods like after the note-taking exercise that will generally be it you know you would complete all your major points but then you would have this really long vocabulary and grammar exercise and that just broke your rhythm that just broke the momentum basically bored you if instead they had some review exercises or review questions that would be much more useful than having those end of you know [?]</p> <hr/> <p>STUDENT A: And there's another point that I'd like to raise I don't know whether they will agree with me or not like the note-taking the lecture from which we had to take notes was so short that our note-taking skills was not actually tested because the questions that followed we could just simply answer them it was such a short time had passed and we could just answer them without consulting our notes so we didn't exactly</p> <hr/> <p>STUDENT C: okay I just realised something I mean TOEFL I give TOEFL [?] probably a lesser challenging level than EASE but TOEFL that boring for some odd reason it had an incentive a motive [?] even the unit of TOEFL that I took first it had an incentive to score high I mean I knew okay I'd score high on it but I wanted to score maximum on it I mean for example I was aiming at 300 out of 300</p> <hr/> <p>STUDENT C: give some scores back it was counted in unit one you scored 59 out of 60 you'd be like motivated okay like next section I'm scoring one hundred per cent</p> <p>STUDENT O (F): I think the exercise and the content of CD two were more interesting and they were better I mean you had to use your mind</p> <p>STUDENT N (F):because they show you the classroom environment and people actually giving presentations</p>
--	---------	---

	Motivation	<p>STUDENT L: I think what makes CD one so boring for us is the fact that we pretty much know what they are going to ask of us there is no level of surprise but obviously monotony sets in because you are like I've already done this stuff before I already know this [?]</p> <hr/> <p>STUDENT H: I think like unit one and two were at one level then slight increase in units three and four and then slight increase in unit five and six and by the time you'd reached unit five and six you were so like tired and just bored out of your mind that you didn't want to do the [?]</p> <hr/> <p>STUDENT M (F): no some stories are funny that old man bio teacher STUDENT O (F): with the fish STUDENT M (F): the fish guy and the STUDENT N (F): the math guy's good</p> <hr/> <p>STUDENT M (F): because the guy was interesting what he was talking about was interesting and he wasn't getting too technical about it and the whole thing on labour markets we're doing like economics they were talking about economics related stuff but it was so very boring</p> <hr/> <p>STUDENT M (F): no not necessarily there was this whole thing on AIDS that was really interesting but we're not studying AIDS I found it really interesting I was able to take notes</p>
Improving Competence	Usability	<p>STUDENT F (F): the logical structure is fine STUDENT B: the transition is very smooth from one section to another that's not a problem STUDENT F (F): you don't see that there's a big like a jump in the middle in the UNKNOWN: A very user-friendly interface STUDENT B: But I thought at times that if there were a little bit of sound effects with the exercises it would have helped because they were really boring at times I'm sorry but it was really boring at times STUDENT F (F): but it becomes a bit monotonous the units can become monotonous because the [?] throughout the CD that become a bit monotonous so if there was a bit more interchanging in the middle that would be good</p> <p>STUDENT H:: and yes that also I was thinking that it had a little glitch in the program because I think they have actually made you save the scrapbook first then whenever you add something you it automatically saved because what happens is that you save it at the end and sometimes our computers would just glitch and they'd shut down and all your notes would be gone</p> <p>STUDENT N (F): yeah my scrapbook got lost a couple of times</p> <p>STUDENT H:: and that's just a glitch in the program</p>

	Suggestions	<p>STUDENT K:: very brief at times they didn't describe the full exercises like I didn't understand the thesis questions which was I didn't understand the questions the statements or how did it explain the exercise that was to follow so I thought that there should be more instructions for those people who are coming from backgrounds like foreign places you know who wouldn't have background</p> <p>STUDENT B: But I thought at times that if there were a little bit of sound effects with the exercises it would have helped because they were really boring at times I'm sorry but it was really boring at times</p> <p>STUDENT F (F): but it becomes a bit monotonous the units can become monotonous because the [?] throughout the CD that become a bit monotonous so if there was a bit more interchanging in the middle that would be good</p> <p>STUDENT B: I think after the note-taking exercise okay basically through the sequential order I think there should be note-taking should be moved a little bit up because basically we see that clip a zillion times before we actually come to the note-taking thing so if we do the note-taking exercise in the middle of it then I think to an extent we'll be able to answer the rest of the exercises with just viewing the tape once and then obviously we'll be able to manage time more effectively and do it more comprehensively than we did I think</p>
	Motivation	<p>STUDENT F (F): but it becomes a bit monotonous the units can become monotonous because the [?] throughout the CD that become a bit monotonous so if there was a bit more interchanging in the middle that would be good</p>
Affective Dimensions	Management	<p>STUDENT C: give some scores back it was counted in unit one you scored 59 out of 60 you'd be like motivated okay like next section I'm scoring one hundred per cent</p> <hr/> <p>STUDENT F (F): I mean he just likes the lectures and he wants to go to Warwick end of story no I think that the topics that were discussed were very interesting and the software itself is really good the first CD I didn't feel as in I found it good but I didn't feel that I learnt so much from it I don't know maybe it was my mind set I'm not saying you know that could be one possibility but I didn't learnt so much from it the second CD I thought I learned from more the presentation because everything we discussed before you know since we are at that point right now that we have to do presentations</p> <hr/> <p>STUDENT A: because we were a lot more interested by default because we have an upcoming presentation in a few days so we all wanted to learn like what new can we do our presentations</p>

	Justification	<p>STUDENT B: But if all of us are complaining about being boring and being dry then if do you expect that the noting exercise is actually 30 minutes long would you be entertained by the lecturer</p> <p>STUDENT L: no but the thing is it teaches you exactly what content is necessary you know when you talk about listening to lectures like I mean the first two units I have done on CD two already it says you know find the appropriate material on the site so basically CD one teaches you how to grasp the essence of the lectures exactly what the person is saying so yeah to an extent it is important as far as improvement is concerned well to an extent yes I guess perhaps I mean unit six teaches antithesis and that taught us something but on the whole not really the level of progress I would have expected</p> <p>STUDENT H: the first CD it was basically skills we had already learnt in the A-levels and O-levels and all so I mean we didn't find it as challenging or we didn't learn as much as we could CD two on the other hand I mean basically no-one's really taught us how to do a presentation properly so I mean [?] you're like oh this is what you do you know this is how you do it very detailed information</p> <hr/> <p>STUDENT J: alright look there's absolutely no pressure on doing that CD you don't have any pressure you don't have it's just a regular class right there's no pressure there's no exam coming up right so it's a kind of relief that alright yeah there was something funny in it so now it's got my attention but if I'd had an exam tomorrow night or in the next month it's a pretty long exam and obviously I don't have time</p> <p>STUDENT O (F): they're not teaching you humour that's a different [?]</p> <p>STUDENT J: that's not what I meant it's basically aimed at keeping your attention about something and it helps but if it comes out unexpectedly then okay but if I have an exam coming up then I wouldn't go for something like that</p> <p>STUDENT H: it would be like that right now if you're having classes and you have an exam tomorrow and the tutor starts cracking a joke and it's like please get on with it</p> <hr/> <p>STUDENT C: as a matter of fact I've actually told my brother to get this EASE CD and do it and I asked him to do it because it's pretty good and his English background is probably not as good these guys but the point is here I really want to do it and I would really have liked if they had the marking scheme so I would have had evaluated how well is he doing but the nonetheless it's worthwhile</p>
--	---------------	--

	Motivation	<p>STUDENT L: I think the fact that you're going to get yourself into university is motivation enough I seriously don't think you need anymore incentive than that.</p> <p>STUDENT G: it was really it was really sad that we couldn't hear the whole lecture like some of the really interesting ones i wish i could just have the whole video i just want to listen to the whole lecture it's just the whole learning they just got chopped off too soon</p> <p>STUDENT F (F): I mean he just likes the lectures and he wants to go to Warwick end of story no I think that the topics that were discussed were very interesting and the software itself is really good the first CD I didn't feel as in I found it good but I didn't feel that I learnt so much from it I don't know maybe it was my mind set I'm not saying you know that could be one possibility but I didn't learnt so much from it the second CD I thought I learned from more the presentation because everything we discussed before you know since we are at that point right now that we have to do presentations</p> <p>STUDENT O and students</p> <p>STUDENT N (F): and students</p> <p>STUDENT M(f): you can relate to them I was like I know what it's like to stand up there and not know what to say</p>
--	------------	--

APPENDIX 4.7 COMMENTS FROM SECTION 4.4.3

126:STUDENT B: at times the vocabulary exercises were annoying as well because there were some sophisticated terminologies used in the lecture which obviously at least I didn't know the spellings of them so whenever I typed them it was just incorrect and it would be an E or an A missing that's a big flaw and even if you put a full stop at the end it puts that it's incorrect

127:STUDENT C: yes in fact a space I mean instead of a single space being two words for example there if you mistakenly press two spaces that's something wrong with the programming of it actually

128:STUDENT A: yeah I would call that a logical error in the programming of the EASE

129:STUDENT G: when it requires you to type in an answer you may type in the right answer but you might put a full stop at the end or a space before or you might be missing a vowel and your answer will be

130:STUDENT F (f): rejected

131:STUDENT G: yeah exactly so it's annoying that's all it's just

132:STUDENT F (f): because in your brain you've given the right answer and you can't really find and you're looking at it and you're like what did I do wrong so there should be some way to pinpoint that okay you know the spelling's wrong or something like that

133:STUDENT C: by the way I figured this out a little bit later if you don't type in an answer for example if press the done button twice it will give you the answer just like that I mean sometimes I didn't realise it was wrong so I just pressed the done button twice and it gave me the answer

TURNS no:108-114

108:STUDENT F (F): the logical structure is fine

109:STUDENT B: the transition is very smooth from one section to another that's not a problem

110:STUDENT F (F): you don't see that there's a big like a jump in the middle in the

111:UNKNOWN: A very user-friendly interface

112:STUDENT B: But I thought at times that if there were a little bit of sound effects with the exercises it would have helped because they were really boring at times I'm sorry but it was really boring at times

113:STUDENT F (F): but it becomes a bit monotonous the units can become monotonous because the [?] throughout the CD that become a bit monotonous so if there was a bit more interchanging in the middle that would be good

114:STUDENT E: And there were a lot of slow periods like after the note-taking exercise that will generally be it you know you would complete all your major points but then you would have this really long vocabulary and grammar exercise and that just broke your rhythm that just broke the momentum basically bored you if instead they had some review exercises or review questions that would be much more useful than having those end of you know [?] and grammar

STUDENT B: at times the vocabulary exercises were annoying as well because there were some sophisticated terminologies used in the lecture which obviously at least I didn't know the spellings of them so whenever I typed them it was just incorrect and it would be an E or an A missing that's a big flaw and even if you put a full stop at the end it puts that it's incorrect

STUDENT C: yes in fact a space I mean instead of a single space being two words for example there if you mistakenly press two spaces that's something wrong with the programming of it actually

STUDENT A: yeah I would call that a logical error in the programming of the EASE

STUDENT G: when it requires you to type in an answer you may type in the right answer but you might put a full stop at the end or a space before or you might be missing a vowel and your answer will be

STUDENT F (F): rejected

STUDENT G: yeah exactly so it's annoying that's all it's just

STUDENT F (F): because in your brain you've given the right answer and you can't really find and you're looking at it and you're like what did I do wrong so there should be some way to pinpoint that okay you know the spelling's wrong or something like that

STUDENT C: by the way I figured this out a little bit later if you don't type in an answer for example if press the done button twice it will give you the answer just like that I mean sometimes I didn't realise it was wrong so I just pressed the done button twice and it gave me the answer

17:STUDENT G: even in an intellectually homogenous university like LUMS where the academics and the instructors and the students are primarily Pakistani you can't escape the fact that some people are from different areas geographically and they've been through different systems so even here we experience a vast difference in accents and pronunciations and something we've had to grapple with coming to Lahore from Karachi so I think it's very very apt even though no material can completely encompass the need for but I mean you need to highlight it it's definitely an issue

81:STUDENT G: for our needs it depends on the audience EASE would be brilliant for an audience which is not doesn't have the same perspectives with the learning experiences as us I mean it did help us but by bringing out the subconscious things to our attention and by doing that definitely I'd never thought of many many many things before it I did the program but then these things were I think they were already subconscious and you knowSTUDENT G: for our needs it depends on the audience EASE would be brilliant for an audience which is not doesn't have the same perspectives with the learning experiences as us I mean it did help us but by bringing out the subconscious things to our attention and by doing that definitely I'd never thought of many many many things before it I did the program but then these things were I think they were already subconscious and you know

83:STUDENT B: okay basically I think there are a lot of things that we take for granted that we assume that we tend to know but despite this I think EASE to an extent obviously it brushed up our listening skills but I thought lecture organisation a lot of critical points that we tend to overlook were actually we were actually forced to actually look at that you were forced to see where the lecturer gave pauses where the lecturer was

laying stress on where he was formal informative whatever different aspects of the lecture were given importance so that essence I think goes on to say a lot about EASE because EASE we might think it's very easy but I think on a subconscious level we've improved a lot and there are a lot of things that we just took for granted that are just coming to light now

91:STUDENT G: I think I can shed some more light on this there's an Eastern saying which says when somebody delivers a speech and a hundred men listen to it each man walks out with his own understanding so basically when you look at it from the lecturer's point of view and you try and analyse it and break it down as you hear it you're probably going to walk out closer to what the lecturer's perspective is and that should and the who student walks out on the same page as the lecturer will be the best of like .

STUDENT B said at some point we might think a peripheral point was a main point but if we had noticed the pauses and the organisation the peripheral we might be giving emphasis to a point which a lecturer didn't think was of primary importance so thinking like the lecturer would help us be on the same page as him

STUDENT F (F): but it becomes a bit monotonous the units can become monotonous because the [?] throughout the CD that become a bit monotonous so if there was a bit more interchanging in the middle that would be good

STUDENT E: And there were a lot of slow periods like after the note-taking exercise that will generally be it you know you would complete all your major points but then you would have this really long vocabulary and grammar exercise and that just broke your rhythm that just broke the momentum basically bored you if instead they had some review exercises or review questions that would be much more useful than having those end of you know [?] and grammar

STUDENT E: [?] I actually liked the visual format of the first CD better because it was light colours it was relaxing and yes it was a bit boring but look within LUMS most of our lecturers are boring it's traditionally it's a fact of life most lecturers are a bit dry they're a bit tedious and those are the most challenging to follow so i think you know CD one was a more realistic selection you want to have a realistic selection of lectures

STUDENT C: as a matter of fact I've actually told my brother to get this EASE CD and do it and I asked him to do it because it's pretty good and his English background is probably not as good these guys but the point is here I really want to do it and I would really have liked if they had the marking scheme so I would have had evaluated how well is he doing but the nonetheless it's worthwhile

APPENDIX 5

APPENDIX 5.1 COMMENTS IN FULL FROM SECTION 5.2.1

5.2.1.1.1 Appropriateness

Positive Comments

4:I think the introduction of EASE was very relevant and the use of both writing that is text and listening will really help me plus the exercises seemed to be very appropriate as answering questions right after the clip is a very is very(sic) helpful for me... (Session One Appropriateness Student RP 1M)

30...overall EASE appears to be a good attempt which I think will help most students overcome their awe for the English language and public speaking...(Session One, Node Appropriateness Student RP8M)

178:The best part about this CD is that it's got viewpoints from students as well as lecturers you know we tend to relate to people from our age better so if you have a student saying something on screen you listen to them more carefully and you give them more attention and give more weight to the points they are telling you rather than a lecturer because you tend to think of lecturers oh you know they're just dictating terms to us another good thing is that over here they're not put in that unnatural classroom environment where everybody had to be formal and everybody's playing a role of a teacher or a student over here it's more informal like the teachers are speaking their parts and so are the students and you're getting some very good points and advice this is all really genuine and it's really good in that way. (WAV Session 7and8 paragraph 178: Student RP20F)

38...we were first made to distinguish between seminars and lectures once this was established we were informed about the usefulness of seminars and how to prepare for them good and bad habits were also mentioned this CD doesn't just make you learn through exercises it provides you with clear and vital information and advice ... (WAV Session7/89. Paragraph 38 Student RP30M)

26...the exercises like the one taking notes where we have to take notes those were pretty helpful and also the ones in which we had to identify the proper use of vocabulary by the instructor so that's pretty good... (Session One, Node Appropriateness Student RP2M)

42...the note taking was a very interesting aspect of the whole thing because none of us knows how to take proper notes and listen to a lecture and make notes at the same time...(Session One, Node Appropriateness paragraph 42 Student RP20F)

30:...with the help of EASE I am learning a lot about evaluation of instructors their attitudes their styles I also learnt about the difference between figurative and literal meanings and when an instructor is using figurative meaning and and (sic) when an instructor is using literal meaning today I also learnt when an instructor is making an argument and I also came to know about two types of argumentation which are thesis antithesis and synthesis I also learnt how arguments are structured...(WAV session 7/8 Student RP7F)

46:Today I finished my fifth unit and due to this now I'm able to evaluate whether the lecturer's claim is strong or less weak now I can easily differentiate in lecture style the manner whether he is serious or comical now I can make judgments about the lecturer's attitude and the most important thing is this that I can now recognise the importance markers I also did some work on nouns adjectives and now I know the literal and figurative meanings as well of the lecture I feel more confident about distinguishing between the significant and the less significant...(WAV Session 7/8 [Cd1] Node Appropriateness Student RP47M).

50...I have just completed the note taking exercises which proved to be extremely helpful though it had little to do with the chapters it was still very effective the speaker was going relatively fast and I had to train myself to go at his pace and not repeat the lecture again and again because I can't stop press stop and rewind in class...
(WAV Session 5 [Cd1] Student RP38F)

78: I think Unit Three was probably the most practical and the most helpful of the units so far because unlike Unit One and Unit Two which were which were quite general Unit Four or Unit Three was specific especially with how instructors use examples in comparing and contrasting in their lectures to help us understand better so this unit was probably the best for far also I did not catch myself dozing off this class because the session was a lot shorter than the other sessions before this so overall it is getting better I guess... (WAV session 4 [Cd1] Node Appropriateness Student RP25M)

58: ...judging the tone of the lecture makes a whole world of difference because occasionally they are used in another context of meaning I think with some further practice rather than getting lost in the speaker's words I shall be able to concentrate more deeply and analyse not only the verbal account but also the speaker's body language facial expressions and tone... (WAV session 2 CD1 Node Appropriateness Student RP8M)

6: I started from Unit number three of EASE 2 seminar skills from this unit I came to know about different types of [?] I also came to know [?] when they are required this unit was very helpful as it also provided information about mistakes students make during their presentations after looking at those mistakes I can now think of the ways to avoid those mistakes in the presentations which count a lot in my academic career it also gave me a view about the advantages and disadvantages of using the computers and presentation software...
(9WAV session Nine CD2/3 node Appropriateness Student RP14M)

297: *the entire EASE CD was really innovating(sic) and the interface used was user friendly I think EASE and other multimedia electronical(sic) education programs are the future of learning and teaching and of course success my vocabulary my communication skills my writing skills the note taking skills are really have been really enhanced after using the EASE I think LUMS and other universities must incorporate EASE and other related programs (WAV session 7/8 CD1/5 Node App: Student RP24M).*

329: *Well this is a pretty comprehensive unit because apart from the examples of presentations there are suggestions from the tutors the differences between a good and a bad presentation and they have simply outlined good and bad student practices which practices that we are not supposed to do in a presentation moreover they have simply outlined different types of presentations presentations which are meant to be for academic purposes or otherwise the role of tutor and student was very clearly broad lined because there are a lot of situations where the student is unable to draw the simple outline between what the teacher is supposed to help what they are supposed to do and what the student is supposed to do moreover the difference between the lecture and the seminar was very good because before that I used to think of a seminar as the same thing as a lecture. (WAV SESSION 7/8 CD 2/1 StudentRP 37M Node App.)*

Negative Comments

...the whole exercise recognising accents was totally baseless and pointless... (WAV Session 2 [CD1/2] Node Appropriateness Student RP 7F)

30: *I didn't find any purpose in having the exercise about backward and forward markers because I don't think it helps us concentrating on helps us in improving our concentration powers... (WAV session 2 [CD1/2 Node app Student RP7F).*

98: *I don't get how we as students are supposed to apply this in a lecture room okay fine we know what transitional statements are but how will it help us in a lecture okay we know the teacher's going on to another point and that we start taking down point number two and point number three in our notebooks but how does that improve our listening skills and the ability to concentrate more on lectures (WAV session 2 CD1/2 Node App Student RP20F)*

248 ...the vocabulary exercises were far too easy I didn't understand the purpose of the vocabulary exercises in which we have to list down the nouns adjectives et cetera and the words even I couldn't figure out what that has got to do with communication skills because that primarily relates to grammar and grammar has nothing to do with communication skills and the exercises relating to figuring out which countries the professors came from were quite irrelevant and stupid discovering which country a person is from by paying attention to his accent doesn't improve one's listening skills but has got to do more with the personal knowledge rather than his listening abilities...(WAV Session 2 [CD1/2] student RP46 M)

70: The second unit I've completed as well I think it was well constructed and the content of language was good but according to me it was overly repetitive like the basic thing that they were focusing on in the first unit was that what if you screw up in the presentation and how that you overcome them and how you use of papers less but I guess you have to overcome them in different way but it was too much like repeating and everything so it was [?] it was boring in the end...(WAV session 7/8 Node Appropriateness Paragraph 70 Student RP 17M Cd2/2)

11:...the exercise on separating literal and figurative phrases was also an easy one I mean it was pretty apparent whether the instructor actually meant what he was saying or where he was not apart from this the exercises in which we were supposed to identify the nouns and adjectives derived from the verbs and their meanings were totally pointless and so were the exercises that required us to fill missing prepositions this I think in no way improves our lecture communication or listening skills it was totally irrelevant this was more of an English grammar exercise and anyone with a sound base in English grammar could easily attempt these exercises with a little help from the clips itself the last few questions of Unit 4 were easy ones as well separating academic and formal language and recognising intensifiers though the distinction between formal and academic language was important as it helps one in using more appropriate words in the notes the exercise on intensifiers although it was easy but more or less it was pointless I think intensifiers are very commonly used and sometimes even unintentionally unless of course instructors take special care before uttering an intensifier. (WAV session 7/8 CD1/unit 4/5 Node Appropriateness Para11Student RP7F)

58: In the vocabulary section of Unit Three I didn't get the point of one of the exercises where you had to see the lecture and see where the lecture uses alphabets from the English language maybe there should be little notes at the beginning or end of each exercise that explained what it is aiming to achieve because frankly I didn't see the point of this one exercise it increases your attention skills and see how attentive you are but I don't know how it helps out in listening to a lecture... (WAV Session FOUR [Cd1] paragraph 58 Node Appropriateness Student RP20F)

5: but I feel if that was the sole purpose to teach us how to dissect arguments then its not, I don't think its focused enough, I don't think it comprehensive enough, to deal with that, but if the topic of, if the focus of this unit is just to # initiate us a or to familiarize us with the # methodology of arguments, how to make an argument how to break an argument and how to analyze and argument how to summarize it, if the aim is just to initiate and familiarize then I think, yes, well, this then. it's then, it's not it isn't badly done, but its not as comprehensive as maybe it could have been because the arguments are very central to a student's life and # perhaps they dealt...dealt with more comprehensively. It would have been of more benefit, not that it's not of benefit as it is, but it's basically, a bit on the superficial side it could have been more comprehensive. (CD 1 Unit 6 WAV Student RP 52 M)

Appendix Table 5.1 Comments from Content, Expectation, Reflection and Motivation

Themes	Categories	Sub categories	Comments
Learning Richness	Appropriateness	Positive	See comments above and in section 5.2.1.1 chapter 5
		Negative	See comments above and in section 5.2.1.1 chapter 5

	Content	<p>Difficulty level</p> <p>Coded passages 199.</p> <p>(191 passages coded from WAV session 1-9)(6+2=8 coded passages from WAV Residual files and left out)</p>	<p>58: the academic level of English in it is quite below par for probably students who have given the SAT and are in LUMS at this stage because it is quite easy to handle around and it's all about listening and just answering back simple feedback to it and it's not exactly very hard to cope with and you'd expect any student to do it quite well... (Wav Session 1 Student RP29M)</p> <p>14: I think the questions that were based on the lectures that was at the end of the lecture were too childish and that didn't help me that was very straightforward and that was a reason to get a bit bored. (WAV sess2 CD1 Para 14 Student RP 19M)</p> <p>58: After completing Unit Two and moving through half Unit Three I feel that the bar has most certainly been raised in terms of the difficulty level and interest some challenges in Unit Two required immense concentration and memorisation for which I was totally caught off guard the most difficult section by far was academic English on the vocabulary because that required memorising specific words from speech without any forewarning another challenging topic was the section where I was required to distinguish between transitions in speeches (WAV Session 2 Paragraphs 57 to 58 Student RP 8M)</p> <p>14: I find it difficult to identify the difference between the speakers that from where he or she belongs to... (WAV Session3 Para 14 Node Difficulty Level Student RP14M)</p> <p>155: So far Unit Five is proving to be the most difficult what I have covered so far has been extremely challenging determining the attitudes of speakers is never easy but in this case many of the lecturers covered their doubts very well analysing the level of conviction and ranking them was even more of a challenge and overall I saw it as a good practice (WAV session 6 Student RP11F).</p> <p>66: so far it's been quite simple but I can see that it can start to become more complex and hopefully it will because that's the way to improve myself for it to be hard so you can improve yourself basically (WAV Session 1 Student RP30M)</p>
		<p>Redundancy</p> <p>53 Passages</p>	<p>94: but it was all very fundamental and I feel like we already know all this all these things and you don't need to know them particularly about how one lecture moves onto another point because these are things that we already know from before so I felt it was a bit a few of the parts were a bit futile because they were too easy and too fundamental and it's interesting to learn about the language the teachers use but I don't think that accent recognition is very important because it doesn't matter what accent the teacher is using as long as they're using proper language and that it's comprehensive and the students understand what they're saying...(WAV session 2CD1/2 Student RP20F)</p> <p>162 ...however at times it was quite repetitive and in that sense quite irritating (WAV Session 7-8 CD1 Unit 6. Student RP15M)</p>
		<p>Sequence</p> <p>97 Passages</p>	<p>138: the first unit and the sub-units are arranged in the right order with increasing level of difficulty same is the case with the video clips and the exercises especially the video clips encourage you to work more... (WAV session 1CD1 Student RP21M)</p> <p>293: overall this CD is very user friendly and the units are arranged at an appropriate level with increasing level of difficulties (CD2 Unit 1WAV sess 7-8 Student RP21M)</p>

		<p>Exercises Coded passages 176</p> <p>WAV sessions 1-9 passages 155+ 14 Residual+ 7 other left out</p>	<p>146: the good thing doing the exercises was that once you were done with them you could compare your answers with the correct ones and figure out how wrong or correct you were in interpreting the message that the lecturer was trying to convey... (CD1 WAV sess2 Student RP33M).</p> <p>188: I just feel that the exercises should not insult intelligence of the students and overkill should be avoided by overkill I mean pointless and stupid exercises on vocabulary and grammar which most students here I think know already... (CD1WAV session2 Student RP4M)</p>
		<p>Clarity 5 passages</p>	<p>22:...in the second part we heard from lecturers and students on how to prepare for a presentation and how to give a good presentation the design of the CD is very easy to use the pattern in which we are slowly being taken into the details of seminars and presentations is very easy to follow and understand everything is explained in very simple English but if there are some words that a student may find difficult understand there is the option of using a dictionary which makes it very convenient to look up meanings of difficult words (CD2 Unit 1-WAV Session 7/8 Student RP42M)</p> <p>534: ...so this exercise was important in the sense that one has to be aware of what context the teacher is talking in the examples used were pretty good and gave an insight into the various styles teachers use to elucidate themselves... (CD 1Unit 5WAV Session 7/8Student RP20F)</p>
	<p>Expectations 4 passages</p>	<p>3: Summarizing the unit 4 and 5, I learnt how we should take start in the presentation and how I must communicate my main points to my audience. What type of things I should apply in order to build my confidence in front of the audience, and make the presentation in such a way that my audience should understand my main point and should not be bored. In general these are all skills give us an idea of how to perform well in the professional world and how to give in my future # future professional presentation. (WAV session 10-12 CD2/4-5, Node Expectation Student RP12M)</p> <p>5: On a final note I believe that it was a constructive experience working on the EASE software and I hope it can be helpful for us in future courses.(WAV session 10-12 CD2/5. Node Expectations Student RP15M)</p>	
	<p>Reflection 25 passages</p>	<p>7: Today's session was the last session of EASE Seminar Skills 2. It has been a great experience using EASE software. It provided me this opportunity to learn different techniques and skills about listening writing and preparing myself for presentations. It gave me a chance to record my own voice after listening to different accents or different students instructors. It was a great and enjoyable experience. (WAV session12 CD2/5 Seminar Skills'. Node: Learning Richness/Reflection Student RP14M)</p> <p>3: I am happy because its what I have gained that matters to me because it has # given me knowledge, it has given me, it has given me practice of taking notes, it has is given me practice for, it has, it has taught me to be a better student altogether at the university level. Thank you very much. (WAV session 10-12 Node Learning Richness/Reflection Student RP22M.)</p>	
	<p>Motivation 192 passages</p>	<p>26:...it's getting a bit tedious like tedious in the sense getting a bit monotonous because it's the same exercises over and over again...(WAV session 5 CD1/3 Student RP10F)</p> <p>70:...Unit Four and have enjoyed to the fullest because of it's ever rising difficulty level which I wanted I practiced on identifying when a particular source was being referred to and then I practice on listing the sources another interesting part came in when I had to judge when the evaluation made of the source by the instructor was either positive or it was not at all evaluated was it being recommended was it just being informed or being acknowledged thus this unit is turning out to be more interesting than I thought ... (WAV session 4CD1 /4, Node Learning richness/Motivation Student RP22M)</p>	

	<u>Effective ness</u>	<p>94: I found the program easy to use easy to navigate around with... (WAV session One Node Improving Competence/ Usability/ Effectiveness. Student RP40M)</p> <p>18: ...the dictionary that is in the software made it easy for me to look up the meaning of any difficult words that I encountered... (WAV Session Two Node Improving Competence/ Usability/ Effectiveness. Student RP 32M).</p> <p>22:...the design of the CD is very easy to use the pattern in which we are slowly being taken into the details of seminars and presentations is very easy to follow and understand everything is explained in very simple English but if there are some words that a student may find difficult understand there is the option of using a dictionary which makes it very convenient to look up meanings of difficult words... (Wav Session 7-8, Node Improving Competence/ Usability/ Effectiveness Student RP 42M)</p> <p>42: ...the first part of the CD the good feature about this CD is that the colours [?] on the software are very bright and vibrant look which makes it use pretty interesting... (Wav Session 7-8, Node Improving Competence/ Usability/ Effectiveness Student RP46M)</p> <p>70: The layout is better because when the clip pops up it automatically goes to one side of the page which is good because you've then got the rest of the page to read all the text whereas in the other one the clip used to pop up in the middle of the page and you had to move around and that used to waste a lot of time... (Wav Session 7-8, Node Improving Competence/ Usability/ Effectiveness Student: RP20F).</p>
	<u>Suggestion s</u>	<p>58: In the vocabulary section of Unit Three I didn't get the point of one of the exercises where you had to see the lecture and see where the lecture uses alphabets from the English language maybe there should be little notes at the beginning or end of each exercise that explained what it is aiming to achieve because frankly I didn't see the point of this one exercise it increases your attention skills and see how attentive you are but I don't know how it helps out in listening to a lecture... (WAV Session 4: Student RP20F pc 29 unit 3 (b))</p> <p>49: Having just completed the EASE CD one thing that I've noticed is that most of the speakers speak on the social sciences on the issues of politics sociology culture literature economics et cetera however not many of them speak on mathematical subjects or subjects dealing purely with science by that I mean the natural sciences now for a student like just me who was taking a computer science major I think that should be there because technical lectures differ not only in content but also very much in style and I think a student should have practice [?] in those kind of lectures as well (WAV 'Session Nine-WAV'CD2/5Student:RP4M)</p> <p>54: ...one place where I do think the program could have been improved is that the explanations to the answers could have been a video clip as well and not just text because the text explanations were not enough so a video explanation would have been better and more appropriate as well(WAV session ONE/ Node Improving Competence/ Suggestion. Student RP25M).</p> <p>10:...the software can further be improved by adding more exercises. (Document 'Session Nine-WAV'Student RP46M).</p>
	<u>CD Compariso n</u>	<p>245: ...Today I started EASE seminar skills one I think this CD overall was better than the last one the interface was better and the exercises were a bit more interesting... (WAV session 7/8 CD1and2 Student RP50M).</p> <p>321:... the CD was also better organised in terms of content and exercises it was far more easy to navigate through these exercises than listening to lectures CD the exercises are short and [?] they make one work harder for these for the answer (WAV session 7/8 CD2 Student RP29M).</p>

	Motivation	<p>170: ...there are more clips in this one I noticed which is better because clips just make it more interesting you know rather than just writing and reading all the time... (WAV session 7/8 CD2/Student RP20F)</p> <p>341: ...I thought the organisation and the structure of the CD is very impressive as it is a very very smooth transition between different parts and sections of the CD the instruction and help content given with the sections is really helpful as it allows us to answer the questions really effectively plus the dictionary and the general user features is very user friendly it's very easy to navigate through this the language used is simple and the available dictionary helps our cause a lot... (WAV session 7/8 CD2 Student RP43M)</p> <p>333: ...the instructional format and technical environment of the CD was more or less the same as the last one and it shared common features this was actually really thoughtful of the creators since I was already used to this format so it was easy to use all the features just the scrapbook or the speech bubble...(WAV session 7/8 CD 1and2: Node Improving competence/Motivation. Student: RP7F).</p>
--	------------	---

APPENDIX 5.3 COMMENTS IN FULL FROM SECTION 5.2.1.3

Appendix Table 5.3Comments Affective Dimensions

Themes	Categories	Comments
Affective Dimensions	Management	<p>118: I propose that both softwares be integrated into next year's comms skills class... (WAV sess7/8 Student RP31M).</p> <p>93: After the first thirty minutes of the session I did find my concentration a little diminishing quite a lot I think it would be better if the class could be divided into smaller sessions like of thirty minutes each with a little break in the middle because I did catch myself dozing off on more than one occasion in this class (WAV session 3 Student RP25M(CD1/end Unit 2)</p>

	<p>Justification</p> <p>46: ...there's one aspect where the program lacks because in real life lecturing if you have to answer questions then there's no way that we can replay a lecture and answer the questions... (WAV Session One Student RP22M).</p> <p>74:...Accent of the instructors remained a problem in some of the exercises today most of the questions that I got wrong were due to the teacher's accent however I think it's good to it's good listening to people from all different parts of the world with their different accents these exercises I mean these exercises are surely going to help me at some part and I mean somewhere in life maybe communicating with foreigners or when understanding foreign instructors in some foreign universities perhaps other than the accent everything is going along... (WAV session 5 Student RP13M).</p> <p>82:...EASE 2 is better than EASE 1 which was listening to lectures because it is more useful and more helpful in our academic life because it tells us different it tells us it gives us advice from different people who are more experienced who have gone through this period but EASE 2 would have been of no use if we had not done EASE 1 because EASE 1 taught us all the basics (Wav Session 7/8 Student RP6M)</p> <p>301:...although the sessions have been long and tiring but they have certainly paid off not only my vocabulary or note taking skills or grammar skills are sharpened but in going to the lectures I have learnt about new topics on history on science on social sciences it really has been a learning experience and a good alternative to traditional classroom experience (WAV Session 7/8 Student RP24M)</p>
--	---

	Motivation	<p>10: ...the note taking and vocabulary exercises are very good because in the vocabulary section there was a part on how to infer the meaning of the foreign words the words you are not accustomed to this was very good because it helped me in evaluating the meanings of the foreign words. (WAV session 6 Student RP37M/Recording 13)</p> <p>7: Today's session was the last session of EASE Seminar Skills 2. It has been a great experience using EASE software. It provided me this opportunity to learn different techniques and skills about listening writing and preparing myself for presentations. It gave me a chance to record my own voice after listening to different accents or different students instructors. It was a great and enjoyable experience. (WAV session 10/12 final session Student Rp14M)</p> <p>131: I just ended Unit Five the lectures are quite interesting it basically the unit basically aims at how foreign lecturers deliver their lecture and we had to recognise certain words which signify their attitude or their opinions on the lecture whether they were being serious or they were they angry or ironical so this basically helps us in improving our skills of listening lectures and this will obviously be helpful when we go abroad because we will be already familiar of what we are going to expect there are certain words such as the bourgeoisie perhaps I still can't pronounce it but because of the previous exercise I was able to recognise what he was saying thus the lecture is quite helpful in all regards all respects ... (WAV session 6. Student RP 39M)</p> <p>226: Thesis antithesis and synthesis are three terms I have absolutely memorised by the end of Unit 6 it was a harder and longer unit in terms of the difficulty level of the questions that we were supposed to answer although I thought I was able to find the main point and subject basically the thesis of these lectures with the help of exercises secondly I thought the antithesis exercises were more interesting as they are [?] of the claims made earlier in the lecture this allowed me to actually use my own intellect and to draw a conclusion as to which side to support so in essence the synthesis antithesis and thesis exercises allowed us at least me to use my own creativity to use my own intellect to come to a conclusion and not only rely on the assumptions or the conclusions drawn by the instructor (WAV session 7/8 Student RP 43M).</p> <p>6: ...a few exercises were about discrimination that we face in nowadays societies and on small scales and large scales basically that was pretty good it gave me the chance to express my views and talk about it... (WAV session 7/8 Student RP9F)</p>
--	------------	--

APPENDIX 5.4 COMMENTS IN FULL FROM SECTION 5.2.2

Appendix Table 5.4 Themes and Categories of RRWA

Themes	Categories	Sub Categories	Comments
Learning Richness	Appropriateness		<p><i>The first class of Communication Skills, in which we were required to use the soft EASE, was if I may say so without sounding sorry, was a breath of fresh air. It was different from the mundane lectures I am used to. Some of the skills being taught in the program, were much needed, especially the one on note taking, which I believe is essential for all college students. It highlighted many of the problems students as a whole face when they attend lectures, such as short attention spans, and the inability to write fast. It was an exercise I didn't do that well in, and I hope further use of this program will allow me to hone my note taking abilities. (Email 1 Sender RP11F)</i></p> <p><i>...i'm still trying to figure out the purpose of backward and forward markers because i don't think that they in anyway do improve our concentration powers (email 103: Sender RP7F)</i></p>
	Content	Difficulty level	<p><i>The vocabulary part of unit2 is something which i wasn't expecting to be difficult, however it was. i mainly found the connotations (sic) part of the exercise to be troublesome. where you had to arrange the connotations (sic) into negative connotation and non negative, there could have been some more explanations (sic) with them, and a bit more guidance would be better. the vocabulary on the whole was not that difficult, instead it was easy for a student of university level. Email 87 Student RP49F).</i></p> <p><i>In my opinion, lectures like the one on biology used a lot of sophisticated language and it was hard to understand and comprehend a lot of technical terms. So, it would be helpful if the language was not so technical and we were able to understand it rather easily. (Email 157 Student RP43M)</i></p> <p><i>I had a lot of trouble with the identification of functions chapter. It was pretty hard and throughout the exercise I kept thinking whether in any whether in any future class i would need to know if the teacher is using an example or is comparing a situation. I came to the conclusion that it was not a very useful exercise. (CD 1 Unit 4 Email 333 Student RP38F).</i></p>
		Redundancy	<p><i>Unfortunately, there was also some very easy and irrelevant stuff there. There were too many "mindless" questions, where you had to fill out the exact phrase used by the speaker, word for word. Almost as pointless were the exercises on vocabulary and grammar. I feel that vocabulary is not a problem for most students who have studied in an English medium, and so exercises of word meanings and idiom meanings felt useless. In fact, I fail to see what place such exercises have in a course on listening skills, as word power is better improved in different ways. (Email 110 sender RP4M).</i></p> <p><i>The same subject matter comes up again in the first 2 exercises of unit, making the experience a bit monotonous... (Email 118 Sender RP11M)</i></p>

		Sequence	<p><i>Having finished unit 1 and having progressed to unit 2, I believe that with the progress of each unit, the difficulty level has raised up a notch. I believe this is a very important factor, because the only way to excel is by challenging yourself. (Email 122 Sender RP48M).</i></p> <hr/> <p><i>...the structure is very logical as it raises the level gradually and quite properly. (Email 147 Sender RP29M).</i></p> <hr/> <p><i>Firstly I found the introduction on how an argument is structured and presented in a lecture to be very interesting, yet challengeing. This was great, as (it) blends nicely into the next section, that of identifying the lecturer's thesis. (Email 200 Sender RP30M)</i></p> <hr/> <p><i>The 5th and the 6th unit overall were at a higher level then the previous ones which was expected. (Email 261 Sender RP49M).</i></p> <hr/> <p><i>This Cd was also better organised in terms of contact and excersises . it was able to navigate through these far more easily than the listening to lecture CD. (Email 294 Sender RP29M)</i></p>
		Exercises	<p><i>The rest of the exercises in this unit, were concerned mostly with note taking I felt , ... And as the units have progressed the difficulty level has risen, the lecturers speak faster , the topics at hand are less general . but I have to say , the difficulty level has risen at a manageable rate .(Email248 Sender RP11F).</i></p> <hr/> <p><i>The organization of the exercises however was much better . There was a lot more information given before and after every exercise , and there wasn't a lot of repetition each exercise is different from the previous. (CD 2 Email 321 Sender RP11F)</i></p>
		Clarity	<p><i>...the layout of this unit was however very long and the instructions were not very clear at times or at times they were very brief.. (Email304 Sender RP49M)</i></p> <hr/> <p><i>...moreover the part of the tutors in helping the students come up with the presentations is clearly given and is a very good part of helping students identify what are they supposed to do. (CD 2 Email 309Sender RP 37M)</i></p>
	Expectations		<p><i>I hope to further my vocabulary, and enhance my speaking skills. I hope to expand the ways I know how to present information, through the learning of new skills. This course, I believe is vital for my future, working life, as in all sectors of business, communication is required. (Email 10 sender RP 30M).</i></p> <p><i>I expect to learn how to be a better student at the university altogether. The cd's are expected to be a source of help to me in case I want to continue my studies abroad. (Email 11 Sender RP22M).</i></p>

	Reflection	<p><i>After completing Unit2 and moving through half of Unit 3 I feel that the bar, most certainly, has been raised in terms of difficulty, effort and interest. Some challenges in Unit 2 required immense concentration and memorization, for which I was caught off-guard. However, after using this software for 2 days I also feel more comfortable and think that the software has motivated me to speak freely with people.</i></p> <hr/> <p><i>Next time I hope to get a a lot more questions right because getting accustomed to the format is essential and I feel Unit 3 will be more challenging. Looking forward to it. (Email 117 Sender RP8M)</i></p> <hr/> <p><i>...Made major points so the subsidiary points missing didnt really hinder my performance when I answered the questions. Not as organized as model notes. Still don't use the page dividing technique, or making tables. My style: write everything one after the other. Thus its all in multiple paragraph forms. Need to work on this.(Email 143 Sender RP20F)</i></p>
	Motivation	<p><i>This session was rather stimulating and kept me awake throughout. It was comprehensive and coherent. (Email 100 Sender RP41F).</i></p> <hr/> <p><i>Well, this was a really disappointing session for me, which sapped much of my motivation. ...In fact, I fail to see what place such exercises have in a course on listening skills, as word power is better improved in different ways. (Email 110 Sender RP 4M).</i></p>

Improving Competence	Usability (the functional and ergonomic aspects of the CD-ROMs or HCI factors).	Glitches	<p>However i did face trouble in Unit one in the note taking section. I took down notes on the yellow paper and went ahead with the questions without saving them. when I went back and retook the notes , and then attempted the questions i found that i could not use my keyboard to answer some of the questions which had answer slots. (Email 38 Student RP 10F).</p> <p>Today, again like the last class, i got wrong at places for not putting some words in commas, etc. I got wrong in question for not putting the word "universal" in double commas as i've done here. Its a machine at the end which is teaching us, and therefore can go wrong at such places. (Email 76 Sender RP13M).</p> <p>Then there were some exercises where i got my answers wrong because i entered "We have" instead of "We've". This is because, the answers stored were possibly in the "We've" format. And at other places, i got wrong because i didnt write "ermm" when the instructor took a pause and started with an "ermmmm" sound. So the software fails at such places. (email 131Sender RP13M).</p> <p>The vocabulary exercises were hard and tough to do. I faced alot of problems in it primarily because alot of times i gave the correct answer but there was a slight spacing or spelling error and the entire answer was termed as incorrect. This was very annoying at times but apart from that the vocabulary exercises improved my english grammar and i learned a few new words. (Email:228 Sender RP43M)</p> <p>...however saving the scrap book is more difficult in this. ive managed to gte(sic) my scrap book lost twice in this session. (Email 348 Sender RP10F).</p> <p>...few of the drawbacks are that when u copy sth from the bubble speech option and paste it , the video closes, next the fact that in some excercises when u click done , nothing happens, u dont get the answer. (email 239/364 RP29M).</p>
		Satisfaction	<p>...it was a very good and helpful experience. the software is a user friendly software. it gives all the required instructions which are needed to work through the program. (116 Sender Rp6M).</p> <p>Today's session was probably the most constructive yet (Email 291 Sender RP10F).</p> <hr/> <p>One thing I noticed was that the layout, as in the colours used, were a bit depressing. The first cd in this respect had a more cheerful setting. (email323 Sender RP20F).</p>
		Effectiveness	<p>I have always liked talking dictionaries because i usually remember definitions when people tell them to me verbally and most of the definitons that i heard there i still remember. (email Sender 242 38F).</p> <p>As far as the interface of the software is concerned, i think it was good, easy to navigate and jump from one unit to another, save scraps at different times, or load the scrapbook during any exercise. (251 Sender RP13M).</p>

	Suggestions	<p><i>I also feel that the simplicity of the written text, and the slight complexity(sic) of the lectures shown may confuse the viewer, i would recommend that the text and lecturers presented be of similar difficulty. (35 sender RP30M).</i></p> <hr/> <p><i>Vocab: don't see the point of the exercise where you have to hear where the lecturer uses alphabets. Maybe there should be little notes at the beginning or end of each exercise that explain what it is aiming to achieve because frankly I didnt see the point of this one. (143 Sender RP20F)</i></p> <hr/> <p><i>There is one problem, though. I think the questions should be asked at the end of the clips. Beecause when the questions are displayed while we are listening to the clip, we tend to gt distracted abd concentrate on gtting our answers right instead of the educational purpose of the clip. (300 sender RP38F).</i></p> <hr/> <p><i>However, i think there should be an additional feature through which we can resume from the place or unit where we leave at the end of the session. That would cut any waste of time due to browsing and finding the place where we left. (251 Sender RP13M)</i></p> <hr/> <p><i>I have just completed unit 5 of the seminar skills CD. One thing that I did not agree with in this unit was the statements they gave for formal and informal presentations because at the end they themselves admitted that these advantages and disadvantages can vary from presentation to presentation and from indiviual to indiviual. So i feel that this excercise should be changed (336 Sender RP18F).</i></p> <hr/> <p><i>Ever since i came to university i have seen alot of teachers use Visual aids. Sometimes they are effective others, just a distraction from our usual lectures and unnecessary. I think that this section should have included a detailed discussion of when and when NOT to use visual aid. (359 Sender RP38F).</i></p> <hr/> <p><i>A few reccomendations though, i thought that if the level of exercises is made a little more challenging than that will definitely help. Secondly, if there is some marking scheme or grading criteria attached to the cd that would help motivate the students. (Email 365 Sender R43M)</i></p>
	CD Comparison	<p><i>This CD (2) is more engaging than the other one as theres more to learn where as the other CD was just a series of exercises which didnt really help. This Cd wasnt so boring and i feel it has helped me differentiate between lectures and seminars and see the different approaches towards presentations and seminars. It was also a smoother run .and i have completed two units entirely and a quarter of the third one. (291 Sender RP10F).</i></p>
	Motivation	<p><i>...i think EASE was a refreshingly new and enjoyable experience. i've never tried anything like this before. it was an extremely user friendly program - navigation was really easy. (Email 46Sender RP7F).</i></p> <hr/> <p><i>I started the CD on Seminar skills today and immediately noticed the difference. The format of this CD , and functionality are totally different . Which was a little bit irritating , because after working on EASE Listening to lectures CD for so long , having to adapt to a new style was really difficult , and on a whole I found the previous format easier to work with. (Email 321: Sender RP11F).</i></p> <hr/> <p><i>Both CDs have similar layouts but completely different formats for the excercises which I found refreshing. Since otherwise it could have become monotonous working on them. (Email 339 RP18F).</i></p>

Affective Dimensions	Management	<p><i>I had fun during the class. But i wish there had been more interaction between the students and the teacher and also among students. We hardly got to know our classmates save our groupmembers. and that is only 2 other students. (Email 367RP 38F).</i></p> <p><i>Over all I think this software was designed according to our course as it covered every aspect of the course.I would recommend to introduce this software permanently for this ss122 course. (Email 391 RP5M)</i></p>
	Justification	<p><i>Exercise 53 relating to vocabulary was quite interesting , it helps u develop listening skills , however I do feel that there wasn't much to learn from the exercise . Overall this has been a constant complaint from me regarding vocabulary exercise , but on a whole I'm glad they are included because they develop your listening skills. Most of the exercises require you to listen out for certain words and phrases at times you've heard each lecture so many times you've memorized it , but these exercises make you pay attention even when you may not want to . so over all I'm glad they're included . (139 Sender RP11F)</i></p> <hr/> <p><i>The past few weeks that i had been doing Ease lecture CD, i assumed it to be quite useless. I thought that it went over things that i already knew, but my opinion changed drastically when i compared the notes of other subjects i had taken with the ones after using ease and that before using it. The change was a lot. Not only was i writing a lot more on papaer at a faster rate but my handwriting was also clearer. I also realised that i was listening to the lecturer more carefully and not only that but i paid more attentrion to what people said in everyday conversations.(Email 300 RP38F)</i></p> <hr/> <p><i>...ease2 is better than ease listening to lectures because it is more useful and more helpful in our academic life...but ease2 would have been of no use if we had not done ease1 because ease1 teaches us the basics... (Email 297 SenderRP5M)</i></p> <hr/> <p><i>I have completed the entire second CD. Content wise I found it to be much better than the first CD. But if I hadn't done the first CD than I might not have been able to fully utilize the second CD. Therefore, I feel that it is important to use the CDs in order. (339 Rp18F).</i></p> <hr/> <p><i>Why i might have felt more favorably to this CD, in retrospect, i feel is because maybe my listening skills and surviving boring lectures has improved. the first CD actually helps you to go through this one. (378 RP10F)</i></p> <hr/> <p><i>Today at the end of session 2 I am feeling lucky to have an instructor like Ms. Saima N. Sherazi because if it was not because of her then I would never have got a chance to have a look at the EASE cd's. (Email 99 Sender RP22M)</i></p>

	Motivation	<p><i>Exemplifying: this I found most interesting. This is a very good way for teachers to not only elucidate the point they are trying to make, but as Prof. Jack Cohen does, make them humorous so that it adds a bit of entertainment to an otherwise boring and monotonous lecture. (62 Sender RP 20F)</i></p> <p><i>Not only was this cd(2) relevant, but it had more clips which made it all the more interesting. And the clips were that of students which is something I could relate to. Because even though the teachers had some amazing things to say, one is always more influenced by what one's peers impressions are. It was also good to see that the students were out of the class room setting and in a more comfortable natural setting where they could speak their thoughts.(email 323 RP20F).</i></p> <p><i>Secondly, if there is some marking scheme or grading criteria attached to the cd that would help motivate the students. (Email 365 Rp43M)</i></p> <p><i>I expect to learn how to be a better student at the university altogether. The cd's are expected to be a source of help to me in case I want to continue my studies abroad. Thus, this course is expected to be a comprehensive one, covering all aspects about learning at the university and more importantly being a better person altogether. (Email 11 Sender RP22M).</i></p> <p><i>I believe that it is a very innovative idea. I have never ever used CDs for learning purposes in my life. It is the first time for me that I am going to learn electronically and I am very excited about it. I hope I will learn a lot from it. I think it is brilliant. We will be directly exposed to listening other people and it will help a lot. This is a real practice and I believe it will have better effects than the normal teaching methods.(Email 12 RP44M)</i></p> <p><i>...the idea of using electronic media is a very innovative one, since it's not a conventional method used in this country. I think the whole concept of a teacher talking and dictating to a class loses its novelty after some time. Thus imparting knowledge using this medium ignited some interest in me from the very first moment...firstly, I hope to gain invaluable information after listening to the lectures of professors from Warwick, since it is a well known and well established institution. (16 Sender RP20 F)</i></p> <p><i>As to my motivation level, it is at a steady level. What would make it even MORE motivating would be a shorter class (122 RP49M).</i></p> <p><i>Secondly, if there is some marking scheme or grading criteria attached to the cd that would help motivate the students. (365 Rp43M)</i></p> <p><i>It guided me to prepare more effective presentations by working with slides and animations, intellectual jokes. (395 RP22M).</i></p>
--	------------	---

Email 317 Sender RP44M

Today, I started the EASE second CD regarding presentation skills. It's a very wonderful software. It introduced me to different types of presentations. The inclusion of the views of students as well as the faculty members was a very good idea. It provided with both sides of the thing. The advice by tutors about how to prepare for the presentations? what is a good exercise and what is not? what is recommended? it all was very interesting. The clips of presentations provided the real view of the situation. It was very informative and very interesting and I learnt a lot from it.

Email 280 Sender RP4M

Having just completed section 6, I can truly say that it was one of the more enjoyable sections so far. The speakers were articulate and engaging, and the exercises were generally helpful in deconstructing arguments. The only thing I feel is that there were not enough challenging exercises, and in many

cases you felt that you were just sitting back and letting the lecturer impart his knowledge, without any practical application or demonstration.

I also completed the first two units on seminar skills. The advice on seminars was excellent, as they thoroughly went over the requirements from a student (before, during and after). However, the advice on presentations was somewhat disappointing. In fact, it was somewhat childish and repetitive in nature, i.e. "don't freeze up", "don't panic", and "speak clearly". No doubt important points, but they're ones which everyone already knows. Practical tips would have been more appreciated (and to be fair, there were a few good points such as making note-cards, and practicing in advance)

I'm now coming to the "meat" of Cd 2, which is Unit 3 and 4 which I hope will prove of great help in preparing my final presentation as well.

Appendix 5.2

NVivo revision 2.0.161

Licensee: Saima Sherazi

Project: EASE User: Administrator Date: 1/26/2007 - 7:22:33 PM
NODE LISTING

Nodes in Set: All Tree Nodes

Created: 12/21/2005 - 10:14:37 AM

Modified: 12/21/2005 - 10:14:37 AM

Number of Nodes: 26

- 1 (1) /Learning Richness
- 2 (1 1) /Learning Richness/Expectations
- 3 (1 3) /Learning Richness/Content
- 4 (1 3 1) /Learning Richness/Content/Difficulty level
- 5 (1 3 2) /Learning Richness/Content/Redundancy
- 6 (1 3 3) /Learning Richness/Content/Sequence
- 7 (1 3 4) /Learning Richness/Content/Exercises
- 8 (1 3 5) /Learning Richness/Content/Clarity
- 9 (1 4) /Learning Richness/Motivation
- 10 (1 6) /Learning Richness/Reflection
- 11 (1 7) /Learning Richness/Appropriateness
- 12 (1 7 1) /Learning Richness/Appropriateness/Positive
- 13 (1 7 2) /Learning Richness/Appropriateness/Negative
- 14 (2) /Improving Competence
- 15 (2 1) /Improving Competence/Usability
- 16 (2 1 1) /Improving Competence/Usability/glitches
- 17 (2 1 5) /Improving Competence/Usability/Satisfaction
- 18 (2 1 6) /Improving Competence/Usability/Effectiveness
- 19 (2 2) /Improving Competence/Suggestions
- 20 (2 3) /Improving Competence/Motivation
- 21 (2 5) /Improving Competence/CD Comparison
- 22 (3) /Affective Dimensions
- 23 (3 1) /Affective Dimensions/Management
- 24 (3 2) /Affective Dimensions/Justification
- 25 (3 3) /Affective Dimensions/Motivation
- 26 (4) /Search Results

cases you felt that you were just sitting back and letting the lecturer impart his knowledge, without any practical application or demonstration.

I also completed the first two units on seminar skills. The advice on seminars was excellent, as they thoroughly went over the requirements from a student (before, during and after). However, the advice on presentations was somewhat disappointing. In fact, it was somewhat childish and repetitive in nature, i.e. "don't freeze up", "don't panic", and "speak clearly". No doubt important points, but they're ones which everyone already knows. Practical tips would have been more appreciated (and to be fair, there were a few good points such as making note-cards, and practicing in advance)

I'm now coming to the "meat" of Cd 2, which is Unit 3 and 4 which I hope will prove of great help in preparing my final presentation as well.

Appendix 5.2

NVivo revision 2.0.161

Licensee: Saima Sherazi

Project: EASE User: Administrator Date: 1/26/2007 - 7:22:33 PM

NODE LISTING

Nodes in Set: All Tree Nodes

Created: 12/21/2005 - 10:14:37 AM

Modified: 12/21/2005 - 10:14:37 AM

Number of Nodes: 26

- 1 (1) /Learning Richness
- 2 (1 1) /Learning Richness/Expectations
- 3 (1 3) /Learning Richness/Content
- 4 (1 3 1) /Learning Richness/Content/Difficulty level
- 5 (1 3 2) /Learning Richness/Content/Redundancy
- 6 (1 3 3) /Learning Richness/Content/Sequence
- 7 (1 3 4) /Learning Richness/Content/Exercises
- 8 (1 3 5) /Learning Richness/Content/Clarity
- 9 (1 4) /Learning Richness/Motivation
- 10 (1 6) /Learning Richness/Reflection
- 11 (1 7) /Learning Richness/Appropriateness
- 12 (1 7 1) /Learning Richness/Appropriateness/Positive
- 13 (1 7 2) /Learning Richness/Appropriateness/Negative
- 14 (2) /Improving Competence
- 15 (2 1) /Improving Competence/Usability
- 16 (2 1 1) /Improving Competence/Usability/glitches
- 17 (2 1 5) /Improving Competence/Usability/Satisfaction
- 18 (2 1 6) /Improving Competence/Usability/Effectiveness
- 19 (2 2) /Improving Competence/Suggestions
- 20 (2 3) /Improving Competence/Motivation
- 21 (2 5) /Improving Competence/CD Comparison
- 22 (3) /Affective Dimensions
- 23 (3 1) /Affective Dimensions/Management
- 24 (3 2) /Affective Dimensions/Justification
- 25 (3 3) /Affective Dimensions/Motivation
- 26 (4) /Search Results

APPENDIX 6

APPENDIX 6.1 PLUM PRE PROGRAM USE QUESTIONNAIRE

PLUM Questionnaires

PLUM Pre Program Questionnaire

Pre- Program Questionnaire for EASE evaluation

Date

Program

Student

Course

Evaluator

Hardware

1 What are you hoping to learn from doing this exercise? Please note down the knowledge or skills you would like to improve in the course of this session.

APPENDIX 6.2 PLUM POST PROGRAM USE QUESTIONNAIRE

PLUM Post Program use Questionnaire EASE Evaluation

Date..... Program: **Listening to Lectures**. Student.....

Course..... Evaluator..... Hardware.....

1 Look back at what you wrote for the Pre program Question and note down

(a) what you learned that you hoped to learn:

(b) what you did not learn that you hoped for:

(c) anything you learned that was unexpected:

2 To what extent do you agree with the following descriptions of the program?

1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree, 5 = strongly disagree.

Please circle one

Easy to operate 1 2 3 4 5

Enjoyable to use 1 2 3 4 5

Provides good support for the exercise 1 2 3 4 5

Provides good advice on how to approach the task 1 2 3 4 5

Helps you learn 1 2 3 4 5

Fits well with the rest of the course 1 2 3 4 5

Well worth the time spent on it 1 2 3 4 5

Please add any further comments if you wish

3. This program is meant to help you improve your academic listening skills.

Could you please comment on any improvements you are aware of in:

a) Your knowledge of how lectures are structured:

b) Your note taking skills:

c) Your approach to listening to lectures:

d) In what ways could the program have done more to help you?

4. Would you want to use it again? Please say why, or why not:

Thank you very much for your help. Your comments will be used to improve the program and the way it is used.

APPENDIX 6.3 RELIABILITY TEST OF LISTENING TO LECTURES

Cronbach's Alpha Reliability test of the measuring instrument PLUM post use questionnaire Question

Alpha values of 0.7 and above are normally regarded as acceptable to good (Hair, *et al.*, 2005). In this study the alpha values 0.79 for EASE Vol 1 Listening to Lectures and 0.89 for EASE vol.2 Seminar Skills thus confirming data reliability.

EASE Vol 1 Reliability

***** Method 1 (space saver) was used for this analysis *****

RELIABILITY ANALYSIS - SCALE (ALPHA)				
Item-total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Alpha if Item Deleted
V1	13.9737	12.6209	.4170	.7819
V2	12.7632	11.6991	.5902	.7498
V3	13.2105	12.3329	.6166	.7507
V4	13.5789	11.8179	.5401	.7592
V5	13.2895	11.7248	.6209	.7450
V6	13.1053	12.2589	.3976	.7890
V7	12.9211	11.4801	.5029	.7683

Reliability Coefficients

N of Cases = 38.0

N of Items = 7

Alpha = .7903

APPENDIX 6.4 RELIABILITY TEST OF SEMINAR SKILLS

*Reliability test of the measuring instrument PLUM Post use Question 2.*A Cronbach's alpha score of 0.8917 for question 2 of the PLUM post-use questionnaire for EASE Vol:2 *Seminar Skills* indicates that the students were honest in their answers⁵; >0.7 is normally deemed acceptable to good level internal consistency (Carmines and Zeller 1979).

EASE Vol: 2 Reliability

***** Method 1 (space saver) will be used for this analysis *****

RELIABILITY ANALYSIS - SCALE (ALPHA)

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Alpha if Item Deleted
V1	11.9167	21.2214	.4339	.9022
V2	11.1667	17.5143	.7136	.8740
V3	11.3611	19.7230	.7145	.8750
V4	11.4722	19.0563	.7189	.8728
V5	11.4167	17.9071	.8497	.8567
V6	11.1944	17.8754	.6995	.8753
V7	11.3056	17.9897	.7313	.8706

Reliability Coefficients

N of Cases = 36.0 N of Items = 7

Alpha = .8917

5 V1: Easy to operate

V2: Enjoyable to use

V3: Provides good support for the exercise

V4: Provides good advice on how to approach the task

V5: Helps you learn

V6: Fits well with the rest of the course

V7: Well worth the time spent on it

APPENDIX 6.5 SUMI QUESTIONNAIRE

SOFTWARE USABILITY MEASUREMENT INVENTORY
(SUMI)

Your name

Name of software

Date

NB the information you provide is kept completely confidential, and no information is stored on computer media that could identify you as a person.

This inventory has fifty statements. Please answer every one of them. Against each statement there are three boxes.

You should mark the first box if you generally AGREE with the statement. Mark the central box if you are UNDECIDED, can't make up your mind, or if the statement has no relevance to your software or to your situation. Mark the right box if you generally DISAGREE with the statement.

In marking the left or right box you are not necessarily indicating *strong* agreement or disagreement but just your general feeling most of the time.

AGREE UNDECIDED DISAGREE

Put a ✓ mark in the box of your choice.

		Disagree		
		Undecided		
		Agree		
		↓	↓	↓
1	This software responds too slowly to inputs.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	I would recommend this software to my colleagues.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	The instructions and prompts are helpful.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	The software has at some time stopped unexpectedly.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Learning to operate this software initially is full of problems.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	I sometimes don't know what to do next with this software.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7	I enjoy my sessions with this software.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	I find that the help information given by this software is not very useful.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	If this software stops, it is not easy to restart it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	It takes too long to learn the software commands.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	I sometimes wonder if I'm using the right command.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	Working with this software is satisfying.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	The way that system information is presented is clear and understandable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14	I feel safer if I use only a few familiar commands or operations.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	The software documentation is very informative.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	This software seems to disrupt the way I normally like to arrange my work.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17	Working with this software is mentally stimulating.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18	There is never enough information on the screen when it's needed.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19	I feel in command of this software when I am using it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20	I prefer to stick to the facilities that I know best.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21	I think this software is inconsistent.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22	I would not like to use this software every day.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23	I can understand and act on the information provided by this software.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24	This software is awkward when I want to do something which is not standard.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25	There is too much to read before you can use the software.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			Disagree	
			Undecided	
			Agree	
		↓	↓	↓
26	Tasks can be performed in a straightforward manner using this software.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27	Using this software is frustrating.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28	The software has helped me overcome any problems I have had in using it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29	The speed of this software is fast enough.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30	I keep having to go back to look at the guides.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31	It is obvious that user needs have been fully taken into consideration.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32	There have been times in using this software when I have felt quite tense.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33	The organisation of the menus or information lists seems quite logical.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- | | | | | |
|----|--|--------------------------|--------------------------|--------------------------|
| 34 | The software allows the user to be economic of keystrokes. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 35 | Learning how to use new functions is difficult. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 36 | There are too many steps required to get something to work. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 37 | I think this software has made me have a headache on occasion. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 38 | Error prevention messages are not adequate. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 39 | It is easy to make the software do exactly what you want. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 40 | I will never learn to use all that is offered in this software. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| | | | | |
| 41 | The software hasn't always done what I was expecting. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 42 | The software has a very attractive presentation. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 43 | Either the amount or quality of the help information varies across the system. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 44 | It is relatively easy to move from one part of a task to another. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 45 | It is easy to forget how to do things with this software. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 46 | This software occasionally behaves in a way which can't be understood. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 47 | This software is really very awkward. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 48 | It is easy to see at a glance what the options are at each stage. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 49 | Getting data files in and out of the system is not easy. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 50 | I have to look for assistance most times when I use this software. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Please check you have ticked each item.

Thank you.

APPENDIX 6.6 SUMISCO REPORT

SUMI Scoring Report from SUMISCO 7.38

Time and date of analysis: 11:34:03 on 04-04-2006

Files used in this analysis:

SUMI English (UK) Language Items

SUMI Version 2.1 Scoring Keys

distributions: set 01A

weights: set 01A

population parameters: set 01A

Data file analysed: ease1d2.ASC: day two EASE Vol1: Listening to lectures

Number of users analysed: 41

Profile Analysis

Scale	UF	Ucl	Medn	Lcl	LF
Global	75	61	58	55	42
Efficiency	77	63	59	55	41
Affect	73	56	53	50	28
Helpfulness	73	65	62	59	49
Control	65	57	54	51	41
Learnability	73	65	61	57	49

Note:

The Median is the middle score when the scores are arranged in numerical order. It is the indicative sample statistic for each usability scale.

The Ucl and Lcl are the Upper and Lower Confidence Limits. They represent the limits within which the theoretical true score lies 95% of the time for this sample of users.

The UF and LF are the Upper and Lower Fences. They represent values beyond which it may be plausibly suspected that a user is not responding with the rest of the group: the user may be responding with an outlier.

Individual User Scores

User	Globa	Effic	Affec	Helpf	Contr	Learn	Identifier
1	61	60	54	64	60	60	001
2	61	71	47	65	52	64	002
3	64	63	66	67	50	61	003
4	71	71	65	67	67	64	004 (C)
5	55	65	54	57	53	63	005
6	57	62	43	56	53	61	006
7	66	63	68	64	57	62	007
8	48	57	22	55	56	68	008(A)
9	23	13	41	22	30	23	009(GEHCL)
10	64	64	66	62	68	63	010 (C)
11	56	65	41	60	58	69	011
12	30	28	34	31	32	13	012(GEHCL)
13	52	54	45	67	50	55	013
14	68	68	66	64	66	71	014(C)
15	52	47	53	61	54	56	015
16	54	47	43	60	53	59	016
17	55	62	44	62	42	58	017
18	54	50	58	60	45	50	018
19	65	64	68	71	55	65	019
20	64	58	64	68	60	66	020
21	37	29	49	44	47	34	021(GEHL)
22	65	71	55	67	55	71	022
23	63	59	56	63	55	63	023
24	57	68	42	59	54	61	024
25	66	71	63	65	62	69	025
26	59	54	63	62	49	59	026
27	53	52	50	56	55	60	027

28	61	69	49	64	57	61	028
29	62	63	56	64	56	66	029
30	56	58	37	55	63	59	030
31	58	53	42	65	50	48	031(L)
32	60	59	57	64	49	61	032
33	65	63	57	69	47	57	033
34	35	39	32	27	48	28	034(GEHL)
35	49	56	44	54	41	65	035
36	63	57	66	58	54	71	036
37	53	54	54	62	34	59	037(C)
38	69	69	52	71	68	66	041(C)
39	63	71	52	65	65	58	042
40	50	48	57	53	54	54	043
41	51	39	42	62	41	35	044 (ECL)

Any scores outside the interval formed by the Upper and Lower Fences are potential outliers. The user who produced an outlier is indicated in the right hand column. The initial letter of the scales in which outliers are found are indicated in parentheses.

Item Consensual Analysis

In the following table, the numbers in the row labelled 'Profile' are the observed responses of the actual users to each item.

The numbers in the row labelled 'Expected' are the number of responses expected on the basis of the standardisation database.

The Goodness of Fit between the observed and expected values is summarised using Chi Square, and these statistics are presented on the line below the expected values.

The number at the end of the Goodness of Fit line is the total Chi Square which applies to that item. The greater the value of the total Chi Square, the more likely it is that the obtained values differ from what is expected from the standardisation database.

Each total Chi Square marked with

*** is at least 99.99% certain to be different

** is at least 99% certain to be different

* is at least 95% certain to be different

Total Chi Square values without asterisks are not likely to differ much from the standardisation database.

In this output, the SUMI items which differ most from the standardisation are presented first.

I would not like to use this software every day.			
Item 22	Agree	Undecided	Disagree
Profile	20	13	8
Expected	8.8	6.91	25.29

Chi Sq	14.24	5.38	11.82	31.44***
--------	-------	------	-------	----------

It is easy to forget how to do things with this software.

Item 45	Agree	Undecided	Disagree	
Profile	3	0	38	
Expected	9.89	7.28	23.83	
Chi Sq	4.8	7.28	8.42	20.49***

I sometimes don't know what to do next with this software.

Item 6	Agree	Undecided	Disagree	
Profile	2	5	34	
Expected	13.14	7.08	20.79	
Chi Sq	9.44	0.61	8.4	18.45***

The software has helped me overcome any problems I have had in using it.

Item 28	Agree	Undecided	Disagree	
Profile	23	14	4	
Expected	11.31	20.48	9.21	
Chi Sq	12.09	2.05	2.95	17.09***

This software seems to disrupt the way I normally like to arrange my work.

Item 16	Agree	Undecided	Disagree	
Profile	10	15	16	
Expected	4.0	9.78	27.22	
Chi Sq	9.03	2.78	4.63	16.43***

The way that system information is presented is clear and understandable.

Item 13	Agree	Undecided	Disagree	
Profile	38	1	2	
Expected	25.83	9.01	6.16	
Chi Sq	5.73	7.12	2.81	15.66***

It is obvious that user needs have been fully taken into consideration.

Item 31	Agree	Undecided	Disagree	
Profile	28	10	3	
Expected	16.62	14.8	9.58	
Chi Sq	7.79	1.55	4.52	13.86***

I think this software has made me have a headache on occasions.

Item 37	Agree	Undecided	Disagree	
Profile	19	2	20	
Expected	9.92	8.57	22.51	
Chi Sq	8.31	5.03	0.28	13.62**

The software has at some time stopped unexpectedly.

Item 4	Agree	Undecided	Disagree	
Profile	9	4	28	
Expected	19.23	4.33	17.44	
Chi Sq	5.44	0.03	6.4	11.87**

I feel safer if I use only a few familiar commands or operations.

Item 14	Agree	Undecided	Disagree	
Profile	10	16	15	
Expected	16.56	7.69	16.76	
Chi Sq	2.6	9.0	0.18	11.78**

The instructions and prompts are helpful.

Item 3	Agree	Undecided	Disagree	
Profile	36	2	3	
Expected	25.53	8.73	6.74	
Chi Sq	4.3	5.19	2.07	11.56**

There have been times in using this software when I have felt quite tense.

Item 32	Agree	Undecided	Disagree	
Profile	6	6	29	
Expected	15.51	6.23	19.26	
Chi Sq	5.83	0.01	4.92	10.76**

I find that the help information given by this software is not very useful.

Item 8	Agree	Undecided	Disagree	
Profile	3	10	28	
Expected	9.11	13.17	18.72	
Chi Sq	4.1	0.76	4.6	9.46**

It is easy to see at a glance what the options are at each stage.

Item 48	Agree	Undecided	Disagree	
Profile	32	7	2	
Expected	22.89	9.75	8.36	
Chi Sq	3.63	0.78	4.84	9.25**

The software allows the user to be economic of keystrokes.

Item 34	Agree	Undecided	Disagree	
Profile	16	17	8	
Expected	25.19	10.26	5.55	
Chi Sq	3.35	4.43	1.08	8.86*

I sometimes wonder if I am using the right command.

Item 11	Agree	Undecided	Disagree	
Profile	6	7	28	
Expected	14.56	6.5	19.94	
Chi Sq	5.03	0.04	3.26	8.33*

There are too many steps required to get something to work.

Item 36	Agree	Undecided	Disagree	
Profile 4	4	33		
Expected	8.5	8.16	24.34	
Chi Sq 2.38	2.12	3.08	7.58*	

Learning how to use new functions is difficult.

Item 35	Agree	Undecided	Disagree	
Profile	4	4	33	
Expected	6.3	10.06	24.65	

Chi Sq	0.84	3.65	2.83	7.32*
--------	------	------	------	-------

It is relatively easy to move from one part of a task to another.

Item 44	Agree	Undecided	Disagree
Profile	36	3	2
Expected	28.3	6.33	6.36
Chi Sq	2.09	1.75	2.99 6.84*

This software occasionally behaves in a way which can't be understood.

Item 46	Agree	Undecided	Disagree
Profile	11	5	25
Expected	13.2	10.33	17.47
Chi Sq	0.37	2.75	3.25 6.36*

I will never learn to use all that is offered in this software.

Item 40	Agree	Undecided	Disagree
Profile	7	8	26
Expected	11.61	11.31	18.08
Chi Sq	1.83	0.97	3.47 6.27*

Either the amount or quality of the help information varies across the system.

Item 43	Agree	Undecided	Disagree
Profile	15	14	12
Expected	12.15	21.36	7.48
Chi Sq	0.67	2.54	2.73 5.93

I feel in command of this software when I am using it.

Item 19	Agree	Undecided	Disagree
Profile	31	6	4
Expected	23.29	10.83	6.87
Chi Sq	2.55	2.16	1.2 5.91

The software hasn't always done what I was expecting.

Item 41	Agree	Undecided	Disagree
Profile	12	10	19
Expected	19.06	9.28	12.66
Chi Sq	2.62	0.06	3.17 5.84

Learning to operate this software initially is full of problems.

Item 5	Agree	Undecided	Disagree
Profile	4	5	32
Expected	8.87	7.58	24.55
Chi Sq	2.67	0.88	2.26 5.82

Using this software is frustrating.

Item 27	Agree	Undecided	Disagree
Profile	3	14	24
Expected	7.01	8.77	25.22
Chi Sq	2.29	3.12	0.06 5.47

The organisation of the menus or information lists seems quite logical.

Item 33	Agree	Undecided	Disagree	
Profile	36	3	2	
Expected	29.35	6.6	5.04	
Chi Sq	1.51	1.97	1.84	5.31

The speed of this software is fast enough.

Item 29	Agree	Undecided	Disagree	
Profile	30	3	8	
Expected	22.99	6.84	11.17	
Chi Sq	2.14	2.15	0.9	5.19

I can understand and act on the information provided by this software.

Item 23	Agree	Undecided	Disagree	
Profile	34	2	5	
Expected	29.62	7.52	3.86	
Chi Sq	0.65	4.05	0.34	5.03

The software documentation is very informative.

Item 15	Agree	Undecided	Disagree	
Profile	21	16	4	
Expected	14.29	20.45	6.26	
Chi Sq	3.15	0.97	0.82	4.94

It takes too long to learn the software commands.

Item 10	Agree	Undecided	Disagree	
Profile	3	3	35	
Expected	4.37	7.18	29.45	
Chi Sq	0.43	2.43	1.04	3.9

The software has a very attractive presentation.

Item 42	Agree	Undecided	Disagree
Profile 17	14	10	
Expected	23.09	11.14	6.77
Chi Sq 1.61	0.73	1.54	3.88

This software is awkward when I want to do something which is not standard.

Item 24	Agree	Undecided	Disagree
Profile 7	21	13	
Expected	12.56	16.76	11.68
Chi Sq 2.46	1.07	0.15	3.68

I enjoy my sessions with this software.

Item 7	Agree	Undecided	Disagree
Profile 19	17	5	
Expected	23.8	11.61	5.59
Chi Sq 0.97	2.5	0.06	3.53

If this software stops it is not easy to restart it.

Item 9	Agree	Undecided	Disagree
Profile 6	11	24	
Expected	6.64	15.81	18.55

Chi Sq 0.06 1.46 1.6 3.12

This software responds too slowly to inputs.

Item 1	Agree	Undecided	Disagree
--------	-------	-----------	----------

Profile 4	5	32	
-----------	---	----	--

Expected	7.79	6.26	26.95
----------	------	------	-------

Chi Sq 1.84	0.25	0.95	3.04
-------------	------	------	------

I keep having to go back to look at the guides.

Item 30	Agree	Undecided	Disagree
---------	-------	-----------	----------

Profile 6	6	29	
-----------	---	----	--

Expected	8.16	9.14	23.7
----------	------	------	------

Chi Sq 0.57	1.08	1.19	2.84
-------------	------	------	------

Tasks can be performed in a straight forward manner using this software.

Item 26	Agree	Undecided	Disagree
---------	-------	-----------	----------

Profile	33	4	4
---------	----	---	---

Expected	28.51	7.04	5.45
----------	-------	------	------

Chi Sq	0.71	1.31	0.39	2.41
--------	------	------	------	------

This software is really very awkward.

Item 47	Agree	Undecided	Disagree
---------	-------	-----------	----------

Profile 4	4	33	
-----------	---	----	--

Expected	2.74	7.21	31.05
----------	------	------	-------

Chi Sq 0.58	1.43	0.12	2.13
-------------	------	------	------

I prefer to stick to the facilities that I know best.

Item 20	Agree	Undecided	Disagree
---------	-------	-----------	----------

Profile 19	11	11	
------------	----	----	--

Expected	17.27	8.53	15.2
----------	-------	------	------

Chi Sq 0.17	0.71	1.16	2.05
-------------	------	------	------

I have to look for assistance most times when I use this software.

Item 50	Agree	Undecided	Disagree
---------	-------	-----------	----------

Profile 5	3	33	
-----------	---	----	--

Expected	5.42	5.99	29.59
----------	------	------	-------

Chi Sq 0.03	1.49	0.39	1.92
-------------	------	------	------

There is never enough information on the screen when it's needed.

Item 18	Agree	Undecided	Disagree
---------	-------	-----------	----------

Profile 4	11	26	
-----------	----	----	--

Expected	7.04	9.18	24.78
----------	------	------	-------

Chi Sq 1.31	0.36	0.06	1.74
-------------	------	------	------

Getting data files in and out of the system is not easy.

Item 49	Agree	Undecided	Disagree
---------	-------	-----------	----------

Profile 8	18	15	
-----------	----	----	--

Expected	5.65	16.83	18.52
----------	------	-------	-------

Chi Sq 0.97	0.08	0.67	1.72
-------------	------	------	------

Working with this software is mentally stimulating.

Item 17	Agree	Undecided	Disagree
Profile 20	14	7	
Expected	16.62	14.19	10.19
Chi Sq 0.69	0.0	1.0	1.69

Working with this software is satisfying.

Item 12	Agree	Undecided	Disagree
Profile 26	11	4	
Expected	22.04	13.34	5.62
Chi Sq 0.71	0.41	0.47	1.59

I think this software is inconsistent.

Item 21	Agree	Undecided	Disagree
Profile 3	12	26	
Expected	5.69	10.6	24.72
Chi Sq 1.27	0.19	0.07	1.52

There is too much to read before you can use the software.

Item 25	Agree	Undecided	Disagree
Profile 8	11	22	
Expected	6.6	9.07	25.32
Chi Sq 0.3	0.41	0.44	1.14

It is easy to make the software do exactly what you want.

Item 39	Agree	Undecided	Disagree
Profile 20	12	9	
Expected	17.06	14.12	9.82
Chi Sq 0.51	0.32	0.07	0.89

I would recommend this software to my colleagues.

Item 2	Agree	Undecided	Disagree
Profile 26	9	6	
Expected	24.51	10.97	5.52
Chi Sq 0.09	0.35	0.04	0.49

Error prevention messages are not adequate.

Item 38	Agree	Undecided	Disagree
Profile 10	15	16	
Expected	10.19	16.59	14.22
Chi Sq 0.0	0.15	0.22	0.38

APPENDIX 7

Appendix 7.1 key logger reports PC 53 Key logger Report

<<DURATION>> 38
<<PROC>> <<TIME 12/20/04 17:45:10>> C:\Program Files\Ease\Ease1.exe
<<WND>>

<<DURATION>> 1
<<PROC>> <<TIME 12/20/04 17:45:11>>
<<WND>>

<<DURATION>> 0
<<PROC>> <<TIME 12/20/04 17:45:11>> C:\Program Files\Ease\Ease1.exe
<<WND>> EASE

<<DURATION>> 0
<<PROC>> <<TIME 12/20/04 17:45:11>>
<<WND>>

<<SCR>> <<TIME 12/20/04 17:45:11>> 041220_174511.jpg

<<DURATION>> 7
<<PROC>> <<TIME 12/20/04 17:45:18>> C:\Program Files\Ease\Ease1.exe
<<WND>> EASE: Unit 2/Structure and organisation

<<SCR>> <<TIME 12/20/04 17:45:26>> 041220_174526.jpg

<<SCR>> <<TIME 12/20/04 17:45:41>> 041220_174541.jpg

<<SCR>> <<TIME 12/20/04 17:45:56>> 041220_174556.jpg

<<SCR>> <<TIME 12/20/04 17:46:11>> 041220_174611.jpg

Italian Cinema 19

<<SCR>> <<TIME 12/20/04 17:46:26>> 041220_174626.jpg

45-790

<<SCR>> <<TIME 12/20/04 17:46:41>> 041220_174641.jpg

Course: Title

<<SCR>> <<TIME 12/20/04 17:46:56>> 041220_174656.jpg

ILecture : Introductio

<<SCR>> <<TIME 12/20/04 17:47:11>> 041220_174711.jpg

n to Neo-Realism

<<SCR>> <<TIME 12/20/04 17:47:26>> 041220_174726.jpg

<<SCR>> <<TIME 12/20/04 17:47:41>> 041220_174741.jpg

<<SCR>> <<TIME 12/20/04 17:47:56>> 041220_174756.jpg

<<SCR>> <<TIME 12/20/04 17:48:11>> 041220_174811.jpg

<<SCR>> <<TIME 12/20/04 17:48:26>> 041220_174826.jpg

<<SCR>> <<TIME 12/20/04 17:48:41>> 041220_174841.jpg

<<SCR>> <<TIME 12/20/04 17:48:56>> 041220_174856.jpg

<<SCR>> <<TIME 12/20/04 17:49:11>> 041220_174911.jpg

<<SCR>> <<TIME 12/20/04 17:49:26>> 041220_174926.jpg

<<SCR>> <<TIME 12/20/04 17:49:41>> 041220_174941.jpg
Title : Introduction to N

<<SCR>> <<TIME 12/20/04 17:49:56>> 041220_174956.jpg
eo-Realsim

<<DURATION>> 291

<<PROC>> <<TIME 12/20/04 17:50:09>> C:\Program Files\Ease\Ease1.exe

<<WND>> Clip 23: Professor Richard Dyer

<<SCR>> <<TIME 12/20/04 17:50:11>> 041220_175011.jpg

<<DURATION>> 4

<<PROC>> <<TIME 12/20/04 17:50:13>> C:\Program Files\Ease\Ease1.exe

<<WND>> EASE: Unit 2/Structure and organisation: 29: Note-taking (iii)

<<SCR>> <<TIME 12/20/04 17:50:26>> 041220_175026.jpg
Richard Dyer 3

<<SCR>> <<TIME 12/20/04 17:50:41>> 041220_175041.jpg
Parts: Neo-Realism Popuar cinema

<<SCR>> <<TIME 12/20/04 17:50:56>> 041220_175056.jpg
practices within italisian c

<<SCR>> <<TIME 12/20/04 17:51:11>> 041220_175111.jpg
onemai trying to

<<SCR>> <<TIME 12/20/04 17:51:26>> 041220_175126.jpg
capture the spirit of popular clasesses.

<<SCR>> <<TIME 12/20/04 17:51:41>> 041220_175141.jpg
epics and comedies and melodramsas..2nd

<<SCR>> <<TIME 12/20/04 17:51:56>> 041220_175156.jpg
part type of film making

<<SCR>> <<TIME 12/20/04 17:52:11>> 041220_175211.jpg

<<DURATION>> 122

<<PROC>> <<TIME 12/20/04 17:52:15>> C:\Program Files\Ease\Ease1.exe

<<WND>> Clip 23: Professor Richard Dyer

<<SCR>> <<TIME 12/20/04 17:52:26>> 041220_175226.jpg

<<SCR>> <<TIME 12/20/04 17:52:41>> 041220_175241.jpg

<<DURATION>> 33

<<PROC>> <<TIME 12/20/04 17:52:48>> C:\Program Files\Ease\Ease1.exe

<<WND>> EASE: Unit 2/Structure and organisation: 29: Note-taking (iii)

<<SCR>> <<TIME 12/20/04 17:52:56>> 041220_175256.jpg
about the ppl neo-

<<SCR>> <<TIME 12/20/04 17:53:11>> 041220_175311.jpg
for the ppl epic, comedy and melodrama.

<<SCR>> <<TIME 12/20/04 17:53:26>> 041220_175326.jpg
as u go on distinction becomes more complex. fina

<<SCR>> <<TIME 12/20/04 17:53:41>> 041220_175341.jpg
l part: feleny : script writed in

<<SCR>> <<TIME 12/20/04 17:53:56>> 041220_175356.jpg
ossessionone: comedies biggest b

<<SCR>> <<TIME 12/20/04 17:54:11>> 041220_175411.jpg
lobox office.

<<SCR>> <<TIME 12/20/04 17:54:26>> 041220_175426.jpg
popular jourism cinema

<<SCR>> <<TIME 12/20/04 17:54:41>> 041220_175441.jpg
. vs. neo-raealism course

APPENDIX 7.2 DESCRIPTION OF UNITS

DESCRIPTION OF UNIT 2 of LISTENING TO LECTURES

Description of EASE Listening to Lectures

When the *EASE Vol 1 Listening to Lectures* is opened at the start the program gives us options to select one of the seven units. These units are:

Introduction and overview

Unit 1: Openings

Unit 2: structures and organizations

Unit 3: Functions 1

Unit 4: Functions 2

Unit 5: Significance and attitude

Unit 6: Argumentation.

In Unit 2 there are further 8 modules. Namely

1. Introduction: The structures of lectures
2. Opening and closing summaries
3. Structure and structural phrases
4. Markers
5. Transitions
6. Note taking
7. Vocabulary
8. Conclusion

1. Introduction: The structures of lectures

When you click on this module the structure and organization page is open. There are many options available to the user shown in the table 1. On this page there is a video running. On the end of this video the user is automatically taken to the next page or if a user wants to skip the video the forward button has to be pressed which takes you to the home page.

2. Opening and closing summaries

When you click on this option it takes to lecture summaries. Besides all the common options there are three clips clip 1, clip 2 and clip 3 which user wants to open. After seeing to the clips the user has to drag each title to the appropriate box. Upon clicking on the forward button the user goes on to the next page where he has to select one of the options. After selecting the options the user selects forward button to go to the home page.

3. Structure and structural phrases

On clicking on this option the user is taken to listening for structure page. Here the user after seeing to clip 1 drag the items in the order the lecturer will talk about them, with the first one at the top. As the user goes forward on the structural phrases page he is provided with the same clip again and in reference to it he has to check the boxes. On moving forward clip 2 is provided and some check boxes to be checked seeing the clip. On moving further ahead listening for detail page opens. On this page the user has to fill in the blanks seeing to the clip 2.1. Next is the verb tense page. Here the user has to select one of the options. Next is the listening for structure page. Here there is space provided for the user to scribble the answer viewing the clip 3. On the next page the clip 3 is given again. Next is the listening to structural phrases where the clip 3 is provided again and in the light of this clip check boxes are checked. On pressing the forward button the home page is displayed.

4. Markers

On clicking this option the user is taken to the forward markers page. Here there are two clips, clip 4 and clip 5. On seeing to this the check boxes are to be checked. The next page is the backward markers. The options available are same as that of the forward markers. The next page is the identifying markers. Here the user matches each clip to the reference contained within it, by clicking on a box on the left and then on the right and seeing to the clip 8, 9, 10 and 11. Next is the intensive listening page where the user has to type in the missing marker after seeing to all the 4 clips. On the clicking the forward or home page the home page is displayed with all the options.

5. Transitions

On clicking on this option the user is taken to transitions page where the user after watching the clip 12 decides which of the phrases signal a closure of a topic or part of a lecture and which signal the opening of a new topic or part of a lecture by dragging each phrase into the appropriate box. Next is the rhetorical questions page in which after watching the video the blank is to be filled. Next is the linking page. The next page is the identifying transitions where user gets the chance to identify transitional stages in a lecture after watching the video. Upon clicking the forward or home button the user is taken to the home page.

5. Note taking

On clicking this option the user is taken to note taking page. The user writes his thoughts. Also a clip is provided of which the user has to make notes about. There are also the options of check boxes and choices. There is also an option to agree or disagree. Upon clicking the next or the forward button the user is taken to the home page.

7. Vocabulary

This option takes you to the idioms page. The user needs to select one of the choices for possible answers. Next is the academic English page where the user has to fill in the blank. Next is the synonym page where user has to find another word in the clip is similar in meaning to the word given. Next is the word analysis which is same like any other page with a clip to be played and choices to be selected. There is also an exercise to scribble the words with particular prefix. Upon clicking the next or the forward button the user is taken to the home page.

8. Conclusion

This is the last option. By clicking this option the user is taken to the conclusion page. There is also an option to view all the clips in the unit 2. Upon clicking the next or the forward button the user is taken to the home page.

Options Table

Options	Functionality
---------	---------------

Back	Go to previous page
Forward	Go to next page
Home	Go to the home page
Open a unit	Displays the list of unit you may want to go
Show dictionary	Opens up the dictionary
Copy text to your scrap book	Copies selected text to your scrap book
Show your scrap book	Shows your scrap book
Print current page	Prints the page
Show help	Provides help

VOL1 UNIT 6 Argumentation

Table2. Options common to all modules

In Unit 2 there are further 10 modules. Namely

1. Introduction: What is an argument?
2. Thesis: The lecture as an argument
3. From thesis to antithesis
4. Antithesis
5. The language and structure of an argument
6. Towards synthesis
7. A final question: Who produces the synthesis?
8. Note taking
9. Vocabulary
10. Conclusion

1. Introduction: What is an argument?

When you click on this module the argumentation page is open. There are many options available to the user as shown in the table 1. On this page there is an audio running. On the end of this audio the user is automatically taken to the next page or if a user wants to skip the video the forward button or the homepage button has to be pressed which takes you to the home page.

2. Thesis: The lecture as an argument

When you click on this option it takes to thesis page. Besides all the common options, there are clips 1.1 and 1.2 which user has to open. After seeing to the clips the user has to select one of the choices given. Upon clicking on the forward button the user goes on to the next page where he has to decide whether a given statement is true or false after watching the clip 1.2 again. Upon clicking on the forward button the user goes on to the next page where he has to make notes after watching the clip 1.2 again. After selecting the options the user selects forward button to go to the home page.

3. From thesis to antithesis

On clicking on this option the user is taken to thesis to antithesis page. Here the user after seeing to clip 2.1 has to type notes under the headings in the box given. As the user goes forward he is provided with the same clip again and in reference to it he has to note down possible ways of questioning it on thesis. On moving forward user has to play clip 2.2 which contains the lecturer's comments on the thesis and answer the question given. On pressing the forward or homepage button the home page is displayed.

4. Antithesis

On clicking on this option the user is taken to antithesis page. On this page the user has to think about the differences between documentary and fiction and drag the phrases on the right into the appropriate box. On the next page the user has to select yes or no option. On moving to the next page the user has to listen to clip 3 and make notes on what the lecturer says beneath the headings in the box. On pressing the forward or homepage button the home page is displayed.

5. The language and structure of an argument

On clicking on this option the user is taken to language page. Here the user after seeing the clip 3 has to tick the phrases the lecturer uses. On the next page there are choices to be selected. On the next page which is structure argument page the user has to listen to the video and drag the boxes in the order on the right, placing the first thing the lecturer does at the top. On pressing the forward or homepage button the home page is displayed.

6. Towards synthesis

On clicking on this option the user is taken to towards synthesis page. Here there are four clips namely clip 4a, 4b, 4c and 4d. User listens to each clip and writes at least one sentence summarizing each extract in the notepad given. On the next page the user needs to make some notes, chose a choice and fill in the blank. The forward button takes user to the pause for thought page where the user compares his thoughts to that of author. On pressing the forward or homepage button the home page is displayed.

7. A final question: Who produces the synthesis?

On clicking on this option the user is taken to final questions page. Here the user types his answers in the notepad. On pressing the forward or homepage button the home page is displayed.

8. Note taking, Vocabulary and Conclusion module are same as that of the Unit 2.

Description of EASE Seminar Skills

When the Ease2 seminar skills is opened from the programs menu in start the program gives us options to select one of the five units. As these units moved on in pretty much the same pattern I believe explaining any two of them will be enough. So I will be explaining the functionality of Unit 1 and Unit 5.

Unit 1:

In Unit 1 there are further 10 modules. Namely

1. Seminars and lectures
2. The roles of tutor and the role of student
3. Do seminars exist in sciences
4. Preparing for a seminar
5. The seminar
6. The purpose and significance of seminars
7. Good and bad student practices
8. Advice from seminar leaders
9. Advice from students
10. roundup

Options	Functionality
Exit	Ends
Back	Go to previous page
Forward	Go to next page
Home	Go to the home page
Open a unit	Displays the list of unit you may want to go
Open the video dictionary	Opens up the dictionary
Copy text to your scrap book	Copies selected text to your scrap book
Show your scrap book	Shows your scrap book
Search by key words	Search
Show help	Provides help
Clips	Play clip
Show help	Provides help

Unit 1 VOL 2

The first module is seminars and lectures. In this module along with other modules the user has one extra option, click to read. By clicking this, the passage to be read is displayed. The second option is the role of tutor and student. Amongst all other common option the user has the option to select a choice after listening to the clip 2. The user also needs to tick the appropriate boxes. After this the user moves to the next module which is "do seminar exists in sciences". In this module the user amongst all other common options also has the option of true/false. The user needs to decide about the truth value of the statement after listening to the clip. Next is the preparing for a seminar section. In this module the user has the extra option of tick the boxes options. The user ticks the correct statement after seeing to the clips. The next module is the seminar section. In this section the user has the additional option of making notes, choosing an option and ticking the correct boxes. The user also has the option to check his answers. The next section is the purpose and significance of seminars. In this section the user can makes notes and compare them to that of the authors. The next is the good and bad student practices. In this section the user needs to drag the options in correct column. The user also needs to make notes. The next section is the advice from leaders. In this section the user drags the options, chooses an option and makes notes. The next is the advice from students. Here the user has the option tick the boxes, choose option and fills in the blanks after seeing to the clips. The next section is the round up section. In this section the user has the option to review all of the previous sections.

Unit 5 Vol2

In Unit 5 there are further 7 modules. Namely

1. introduction
2. Referring to sources
3. Qualifications
4. A place for subjects vies and absolute statements
5. Personal and impersonal styles
6. The use of pronouns

7. roundup

Options	Functionality
Exit	Ends
Back	Go to previous page
Forward	Go to next page
Home	Go to the home page
Open a unit	Displays the list of unit you may want to go
Open the video dictionary	Opens up the dictionary
Copy text to your scrap book	Copies selected text to your scrap book
Show your scrap book	Shows your scrap book
Search by key words	Search
Show help	Provides help
Clips	Play clip
Show help	Provides help

Table 1.
common to all modules

Options

The first section is the introduction section. Here the user just goes through the instructions and move on to the next section. The next section is the referring to sources. Here the user has the options of making notes, filling in the blanks, ticking the boxes, listen to model and record yourself besides all common options. The next section is the qualification sections. In this section amongst all other options the user has the option to fill in the blanks and ticking the correct boxes after listening to the clip. The user also has the option to read an excerpt in this section. The next section is the place for subject views and absolute statements. In this section the user also has the option to choose options, ticks the boxes and option to read the excerpts. The next is the personal and impersonal styles section. In this section the user has to make notes and drags the option in the correct boxes. The next section is the use of pronouns section. In this section the user has to make notes and ticks the boxes. The last section is the roundup section. In this section the user has the option to review the previous sections individually.

APPENDIX 7.3 STUDENTS' NAVIGATION ACCOUNTS

Appendix CHAPTER 7 Individual Navigation Accounts

Show in Excel in abbreviated form.

Sections of Unit 2 Vol1	Pc 53	Pc54	Pc56	Pc57	Pc59
1. Introduction: The structures of lectures					
2. Opening and closing summaries					
3. Structure and structural phrases					
4. Markers					
5. Transitions					
6. Note taking					
7. Vocabulary					

8. Conclusion					
---------------	--	--	--	--	--

Unit 2 pc 53

D:\Session2Images(15-12-04)\LAB2PC53\Screens This person has been working on finishing Unit 1 two days data 15and 20/12/04

Started unit 2 at 18:40:17 OPENS page 1/52 The person on this pc jumps directly from home page to the identifying markers page. There s/he clicks the clip 8 option. After watching it s/he matches the labels with the clip. He does the same for the other three clips, clip 9, 10 and 11. Once he is through this he checks for the feedback on his matching. After this he opens his scrap book apparently for no reason. Next he moves on to intensive learning page. On this page the user fills in the blanks after seeing the clip 8 again. Then checks for the feedback. After this he refills the blanks and checks for the answer and this time gets it right. The user does the same for the clip 9, 10 and 11 except for the feedbacks. Next he moves to the transitions page. Over here he watches clip 12 and drags the labels accordingly and checks for the feedback. For clips 13 and 14 he does the same for what he had done for clips 8, 9, 10 and 11. Next he moves to the rhetorical questions page. Here the user sees the clip 15 and fills in the blank. Next he moves to the linking page where the user scribbles on the space provided. Next he opens clip 16 and compares what he has written with the clip. Next is the identifying transitions page where the user watches the clip. On the next page the user watches the clip 19 to 22 and decides the transitions. Next is the note taking section where the user makes notes. The user spends 20 minutes on this unit. Incomplete. Other session

Unit 2 pc 54

The user goes to the introduction page. From there he moves on to the lecture summaries page. Here the user watches the clips and drag the labels under appropriate headings. Then the user listens to the three clips and decides which one was a closing ceremony and checks for the feedback. Next is the listening for lectures. On this page the user listens to the clip 1 again and drags the labels in order. The next is the structural phrases page where the user listens to the clip 1 again and ticks the sentence the lecturer uses. Next is the subject matter page where the user listens to the clip 2 and ticks the things the lectures uses. Next is the listening for detail page. Here the user fills in the blank in reference to the clip 2.1 and then checks for the feedback. Next is the verb tense page where the user needs to select a choice. Next is the listening for structure page where the user listens to the clip and fills in the blank and checks for the feedback. On the next page the user listens to the clip 3 again and drags the labels in the space provided. Next is the listening to structural phrases where the user sees the clip 3 and ticks the write sentences checks for the feedback and asks for some extra advice. Next the user moves to the forward marker page. On this page the user plays the clip 4 and 5 and checks the write markers. He does the same for clip 6 and 7. The user next moves to the identifying markers page. After watching the clips given he matches the labels with the clip. He does the same for the other three clips, clip 9, 10 and 11. Next he moves on to intensive learning page. On this page the user fills in the blanks after seeing the clip 8 again. And does the same for clips 9, 10 and 11. Next he moves to the transitions page. Over here he watches clip 12 and drag the labels. For clips 13 and 14 he does the same for what he had done for clips 8, 9, 10 and 11. Next he moves to the rhetorical questions page. Here the user sees the clip 15 and fills in the blank and checks for the feedback. Next he moves to the linking page where the user scribbles on the space provided. Next he opens clip 16 and compares what he has written with the clip. Next is the identifying transitions page where the user watches the clip. On the next page the user watches the clip 19 to 22 and decides the transitions. Next is the note taking section where the user makes notes seeing to the clip 23. The user spends 50 minutes on this unit.

Unit 2 pc 56

The user goes to the introduction page. From there he moves on to the lecture summaries page. Here the user watches the clips. He skips the next page and moves on. Next is the listening for lectures. On this page the user listens to the clip 1 again. The next is the structural phrases page where the user listens to the clip 1 again and ticks the sentence the lecturer uses. Next is the subject matter page where the user listens to the clip 2 and ticks

the things the lectures uses. Next is the listening for detail page. Here the user fills in the blank without seeing the clip 2.1. Next is the verb tense page. The user skips this page and moves on. Next is the listening for structure page where the user listens to the clip and fills in the blank and checks for the feedback. On the next page the user listens to the clip 3 again and moves on. Next is the listening to structural phrases where the user sees the clip 3 and ticks the write sentences. Next the user moves to the forward marker page. On this page the user plays the clip 4 and 5 and checks the write markers. He does the same for clip 6 and 7. The user next moves to the identifying markers page. After watching the clips given he matches the labels with the clip. He does the same for the other three clips, clip 9, 10 and 11. Next he moves on to intensive learning page. On this page the user fills in the blanks after seeing the clip 8 again. And does the same for clips 9, 10 and 11. Next he moves to the transitions page. Over here he watches clip 12 and drag the labels. For clips 13 and 14 he does nothing and moves on. Next he moves to the rhetorical questions page. Here the user sees the clip 15 and moves on. Next he moves to the linking page where the user scribbles on the space provided. Next he opens his scrap book and copy the selection to it. Next he opens clip 16 and compares what he has written with the clip. He does the same for the clip 17 and 18. Next is the identifying transitions page where the user watches the clip. On the next page the user watches the clip 19 to 22 and decides the transitions. Next is the note taking section where the user makes notes seeing to the clip 23. Next he drags the labels under appropriate headings with the help of scrap book. On the next page the user ticks the statements that are true. The user skips the next page and on the next page selects the option and checks for the feed back. The user next moves to the idioms page. And without selecting any choice moves on to Academic English page. Here the user fills the blanks. The user spends 1 hour and 1minute on this unit.

UNIT 6:

Introduction: What is an argument?
Thesis: The lecture as an argument
From thesis to antithesis
Antithesis
The language and structure of an argument
Towards synthesis
A final question: Who produces the synthesis?
Note taking
Vocabulary
Conclusion

Unit 6 pc 03

The user starts this unit at 18:14. The user goes over the introduction for 3 minutes. He starts the next section which is thesis at 18:17. On the first page of this section the user selects an option without seeing the clip 1.1 provided and checks for the feedback.

On the next page he watches the clip 1.2 and does not select any option and moves on to the next page. On the next page the user watches the clip 1.2 and decides about the truth value of the sentences. The user moves on and on the next page watches the clip 1.2 again and types his answer. The same is done on the next page for clip 1.5. The next page is the pause for thought page. Here the user agrees with statement and comments. Then he checks for the feedback and moves on to the home page. User took about 6 minutes on this section. The user then moves on to from thesis to antithesis page. On this page the user watches the clip 2.1 and types notes under each heading given in the box. Next he moves to the page where he compares his notes with that of the author. On the next page he watches the clip 2.1 again. On the next page he watches the clip 2.2 and selects an option. The user took 7 minutes to complete this unit. The user then moves on to antithesis section. On the first page the user drags the phrases into the appropriate boxes. On the next page the user selects the no option and checks for the feedback. On the next page the user watches the clip 3 and makes notes. On the next the user compares his notes with that of the author. The user took 5 minutes on this section. Next the user moves on to the language of argument section. On the first page the user is provided with clip 3 but ticks the options without having a glance at the clip and moves on. On the next page the user selects two options and moves on. The user next moves on to the structure of argument page. Here he drags the phrases in order after seeing the clip. The user took about 4 minutes on this section. Next the user moves on to the towards synthesis page. Here he watches the clips 4a, 4b, 4c and 4d and makes notes about them. On the next page he compares his results with that of the author. On the next page he opens his scrap book. He took about 10 minutes on this section. Total time 35 mins.

Unit 6 pc 19

The user starts this unit at 18:11. The user goes over the introduction for 5 minutes. He starts the next section which is thesis at 18:16. On the first page of this section the user selects an option after seeing the clip 1.1 provided and opens his scrapbook. On the next page he watches the clip 1.2 and selects one option and moves on to the next page. On the next page the user watches the clip 1.2 and decides about the truth value of the sentences and checks them with the feedback option and copies the text to the scrap book. The user moves on and on the next page watches the clip 1.2 again and types his answer and compares them to that of the authors. The same is done on the next page for clip 1.5. The next page is the pause for thought page. Here the user agrees with statement and comments. Then he checks for the feedback and moves on to the home page. User took about 9 minutes on this section. The user then moves on to the thesis to antithesis page. On this page the user watches the clip 2.1 and types notes under each heading given in the box. Next he moves to the page where he compares his notes with that of the author. On the next page he watches the clip 2.1 again and makes notes. On the next page he watches the clip 2.2 and selects an option after having a look at the scrapbook. Then he checks his answers with the feedback. The user took 15 minutes to complete this unit/section. The user then moves on to antithesis section. On the first page the user drags the phrases into the appropriate boxes and checks for the feedback. On the next page the user selects the no option and checks for the feedback. On the next page the user watches the clip 3 and makes notes. On the next the user compares his notes with that of the author. The user took 5 minutes on this section. Next the user moves on to the language of argument section. On the first page the user is provided with clip 3 and ticks the options after having a glance at the clip, checks for the feedback and moves on. On the next page the user uses feedback options and moves on. The user next moves on to the structure of argument page. Here he drags the phrases in order after seeing the clip and checks for the feedback. The user took about 15 minutes on this section. Next the user moves on to the towards synthesis page. Here he watches the clips 4a, 4b, 4c and 4d and makes notes about them. On the next page he compares his results with that of the author. On the next page he opens dictionary. On the next page the user listens to the clip 4e and fills in the blanks. Next the user moves to the pause for thought page and compares his thoughts with that of the author. Next he moves to the final question page makes notes and checks the author's views. He took about 25 minutes on this section. The user then moves on to the note taking section. Here the user listens to the clip 5 and takes notes. On the next page the user selects one of the five options, checks for the feedbacks and moves on. Next the user ticks some items, checks for the feedback and moves on. Next he moves on to the page where the user drags the statements into the appropriate boxes. On the next few pages the users selects the option and checks for the feedback and

moves. Next the user goes to the explaining terminology page and selects the choices. On the next page the user makes notes. Next the user moves on to the expressions with similar meaning page and fills in the blanks and checks for the feed back. The user did this section in 35 minutes.

Unit 6 pc 20

The user starts this unit at 18:59. The user goes over the introduction for 1 minute. He starts the next section which is thesis at 19:00. On the first page of this section the user selects an option after seeing the clip 1.1 provided and checks for the feed back. On the next page he watches the clip 1.2 and selects one option and moves on to the next page. On the next page the user watches the clip 1.2 and decides about the truth value of the sentences and checks them with the feed back option. The user moves on and on the next page watches the clip 1.2 again and types his answer and compares them to that of the authors. User took about 5 minutes on this section. The user then moves on to the thesis to antithesis page. On this page the user watches the clip 2.1 and types notes under each heading given in the box. Next he moves to the page where he compares his notes with that of the author. On the next page he watches the clip 2.1 again and makes notes. On the next page he watches the clip 2.2 and selects an option and checks for the feedback. Then he checks his answers with the feedback. The user took 8 minutes to complete this unit. The user then moves on to antithesis section. On the first page the user drags the phrases into the appropriate boxes and checks for the feedback. On the next page the user selects the no option and checks for the feedback. On the next page the user watches the clip 3 and makes notes. On the next the user compares his notes with that of the author. The user took 7 minutes on this section. Next the user moves on to the language of argument section. On the first page the user is provided with clip 3 and ticks the options after having a glance at the clip, checks for the feedback and moves on. On the next page the user uses feedback options and moves on. The user next moves on to the structure of argument page. Here he drags the phrases in order after seeing the clip. The user took about 7 minutes on this section. Next the user moves on to the towards synthesis page. Here he watches the clips 4a, 4b, 4c and 4d and makes notes about them. On the next page he compares his results with that of the author. On the next page the user listens to the clip 4e and fills in the blanks. Next the user moves to the pause for thought page and compares his thoughts with that of the author. Next he moves to the final question page and checks the author's views. He took about 6 minutes on this section. The user then moves on to the note taking section. Here the user listens to the clip 5 and takes notes. On the next page the user selects one of the options, checks for the feed backs and moves on. Next the user ticks some items, checks for the feedback and moves on. Next he moves on to the page where the user drags the statements into the appropriate boxes. On the next few pages the users selects the option and checks for the feed back and moves. The user spends 24 minutes on this section. Next the user goes to the vocabulary section. The first page is explaining terminology page and user selects the choices. On the next page the user makes notes. Next the user moves on to the expressions with similar meaning page and fills in the blanks and checks for the feed back. The user did this section in 4 minutes.

Unit 6 pc 21

The user starts this unit at 18:57. The user goes over the introduction and has a look at the scrap book option. The user spent 3 minutes on this section. He starts the next section which is thesis at 19:00. On the first page of this section the user does not select any option after seeing the clip 1.1 provided. On the next page he watches the clip 1.2, selects one option, checks for feed back and moves on to the next page. On the next page the user watches the clip 1.2 and decides about the truth value of the sentences and checks them with the feed back option. The user moves on and on the next page watches the clip 1.2 again and types his answer and compares them to that of the authors. User took about 5 minutes on this section. The user then moves on to the thesis to antithesis page. On this page the user watches the clip 2.1 and types notes under each heading given in the box. Here he also uses dictionary. Next he moves to the page where he compares his notes with that of the author. On the next page he watches the clip 2.1 again and make notes. On the next page he watches the clip 2.2 and selects an option, opens the dictionary and checks for the feedback. The user took 8 minutes to complete this unit. The user then moves on to antithesis section. On the first page the user drags the phrases into the appropriate boxes and checks for the feedback. On the next page the user selects the no option and checks for the feedback. On the next page the user watches the clip 3 and makes notes. On the next the user compares his notes with that of the author. The user took 8 minutes on this section. Next the user moves on to the language of argument section. On the first page the user is provided with clip 3 and ticks the options after having a glance at the clip, checks for the feedback and moves on. On the next page the user uses feedback options and moves

on. The user next moves on to the structure of argument page. Here he drags the phrases in order after seeing the clip. The user took about 3 minutes on this section. Next the user moves on to the towards synthesis page. Here he watches the clips 4a, 4b, 4c and 4d and makes notes about them. Here the user also uses the dictionary. On the next page he compares his results with that of the author. On the next page the user listens to the clip 4e and fills in the blanks. Next the user moves to the pause for thought page and compares his thoughts with that of the author. Next he moves to the final question page and checks the author's views. He took about 11 minutes on this section. The user then moves on to the note taking section. Here the user listens to the clip 5 and takes notes. On the next page the user selects one of the options, checks for the feed backs and moves on. Next the user ticks some items, checks for the feedback and moves on. Next he moves on to the page where the user drags the statements into the appropriate boxes. On the next few pages the users selects the option and checks for the feed back and moves. The user spends 21 minutes on this section. Next the user goes to the vocabulary section. The first page is explaining terminology page and user selects the choices. On the next page the user makes notes. Next the user moves on to the expressions with similar meaning page and fills in the blanks and checks for the feed back. The user did this section in 5 minutes.

Unit 6 pc 24

The user starts and directly jumps to the note taking section. Next the user ticks some items, checks for the feedback and moves on. Next he moves on to the page where the user drags the statements into the appropriate boxes and checks for the feed back. On the next few pages the users selects the option and checks for the feed back and moves. Next the user goes to the explaining terminology page and selects the choices. On the next page the user makes notes. The user did this section in 2 minutes Next the user moves on to the expressions with similar meaning page and fills in the blanks, takes help from the dictionary and checks for the feed back. On the next page the user matches the labels and checks for the feed back. The user did this section in 9 minutes.

SEMINAR SKILLS UNITS 1 and 5

Individual Students navigation through Unit 1 of EASE TWO

Pc27

The user starts the first section at 18:04. Here he drags the option after watching the clip, checks for the feed back, reads the comments and moves on to the next section. He starts the next section at 18:09. Here the user selects one of the options, ticks the correct answer, check for his answers and moves on to the next page. He starts the next section at 18:12. Here the user ticks the correct answers, check for his answer and moves on. The user starts the next section at 18:17. Here he ticks the correct answer after seeing the clip, checks for the answers and moves on. The user starts the next section at 18:18. Here he selects he watches the clip, selects an option, scribbles in the notepad, checks for the answer, compare his views to that of the author and moves on to the next page. The user starts the next section at 18:24. Here the user watches the clip, makes notes, compare his notes and moves on. He then starts the next section at 18:27. Here the user watches a clip, drags the statements, makes notes and moves on. The user starts the next section at 18:38. Here the user watches a clip, drags the statements, selects an option, checks his answers and moves on. The user starts the next section at 18:44. Here the user watches a clip, ticks the correct box, fills in the blanks, checks for the answers and moves on. The user starts with next section at 18:53 and just skims through it.

Pc 30

The user starts the first section at 17:55. Here he drags the option after watching the clip, checks for the feed back, reads the comments, opens scrapbook and moves on to the next section. He starts the next section at 17:59. Here the user selects one of the options, ticks the correct answer, check for his answers, opens the scrap book and moves on to the next page. He starts the next section at 18:01. Here the user ticks the correct answers, check for his answer, open scrap book and moves on. The user starts the next section at 18:03. Here he ticks the correct answer after seeing the clip and moves on. The user starts the next section at 18:05. Here he selects he watches the clip, selects an option, scribbles in the notepad, checks for the answer, compare his views to that of the author and moves on to the next page. The user starts the next section at 18:12. Here the user watches the clip, makes notes, compare his notes and moves on. He then starts the next section at 18:15. Here the user watches a clip, drags the statements, makes notes and moves on. The user starts the next section at 18:24. Here the user watches a clip, drags the statements, selects an option, checks his answers and moves on. The user

starts the next section at 18:29. Here the user watches a clip, ticks the correct box, fills in the blanks, checks for the answers and moves on. The user starts with next section at 18:36 and just skims through it.

Individual Students navigation through Unit 5 of EASE TWO

Pc 27

The user goes to the section 3 directly. He starts the section at 18:40. Here the user watches the clip, ticks the boxes and moves on. He starts the next section at 18:44. Here the user watches the clip, ticks boxes, checks his answers and moves on to the next section. The user starts the next section at 19:00. Here the user watches the clip, drags the statements, makes notes and move on. The user starts the next section at 19:07. Here the user watches the clip, drags the statements, makes notes, checks the answer and moves on. The user starts the next section at 19:11. As this is the roundup section just skims through it.

Pc 30

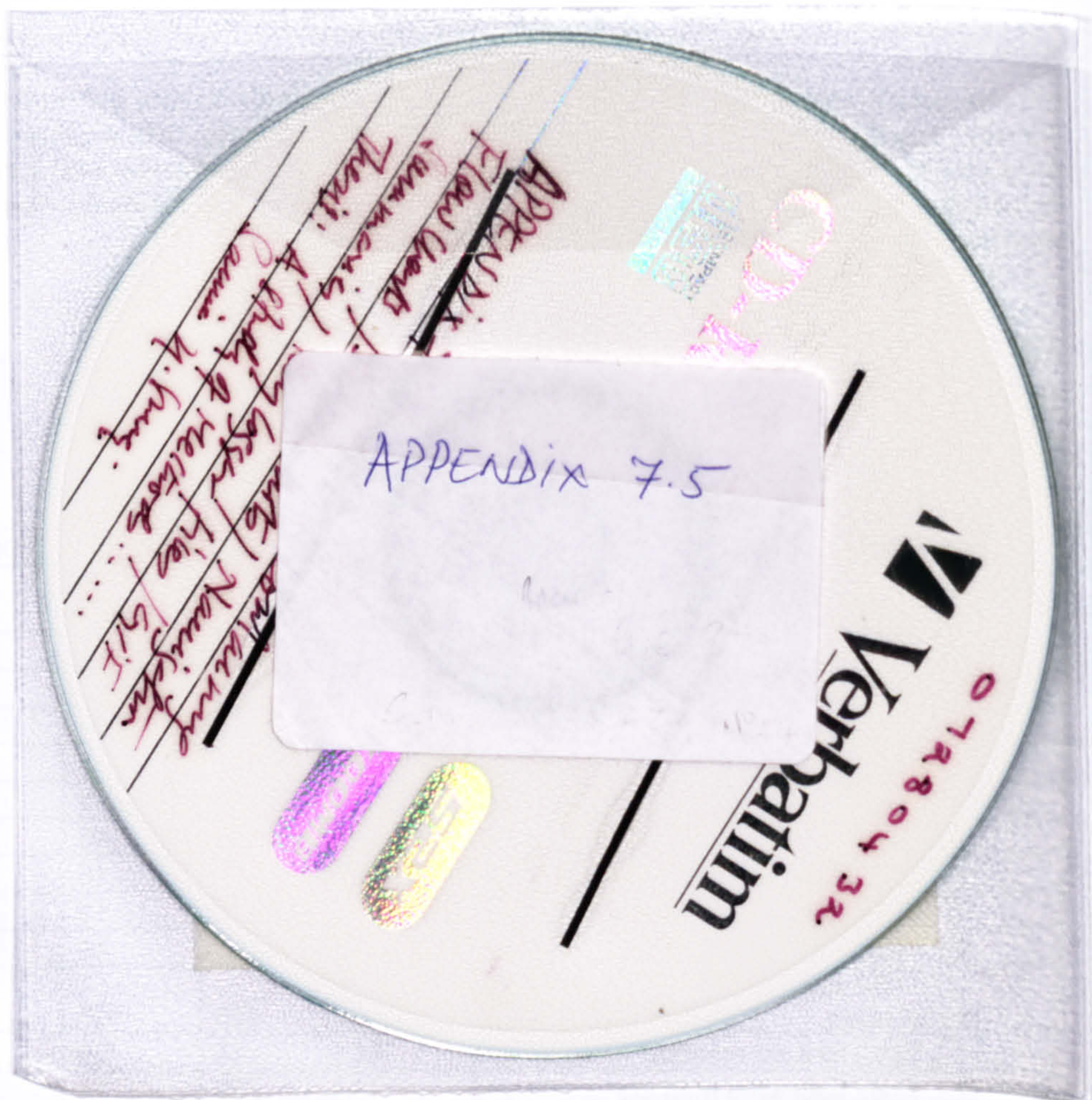
The user starts this section at 19:33:17. Here he just skims through the material and moves on to the next page. The user starts the next section at 19:33:47. Here the user watches the clip, ticks the boxes, makes notes, checks the answer, fills the blanks, opens scrap book and moves on. The user starts the next section at 19:46:57 Here the user opens the video dictionary, watches the clip, ticks the boxes, fills the blanks, checks the answer, opens scrap book and moves on. This is the last section he does. Finishes 18:09:08

APPENDIX 7.4 TIME USAGE WORKSHEET

TIME USAGE of FIVE STUDENTS working on Unit 2 of Ease Vol 1

Excel Worksheet

Section	Page/Name	PC53	PC54	Pc56	PC57	PC62
0	1/52 Intro to Structure and organisation	60	45	30	45	60
1	2/52 lecture Summaries(i)	180	90	120	195	90
2	3/52 lecture summaries (ii)	30	90	15	15	15
3	4/52 Listening fr structure(i)	105	90	75	195	60
4	5/52 structural phrases	165	90	90	105	120
5	6/52 Subject matter	30	45	15	105	30
6	7/52 Listening for detail	75	60	45	60	30
7	8/52 verb tense	75	30	15	30	60
8	9/52 listening for structure(ii)	45	75	15	90	120
9	10/52 listening for structure (iii)	30	45	60	60	15
10	11/52 listening to structural phrases	165	135	105	75	90
11	12/52 forward markers	90	75	60	105	0
12	13/52 backward markers	45	60	75	45	0
13	14/52 identifying markers	270	165	225	180	0
14	15/52 intensive listening (i)	150	45	45	45	0
15	16/52 intensive listening (ii)	60	45	60	45	0
16	17/52 intensive listening (iii)	45	45	15	45	0
17	18/52 intensive listening (iv)	75	45	45	45	0
18	19/52 transitions	90	75	30	135	75
19	20/52 transitions (ii)	75	90	30	60	75
20	21/52 transitions (iii)	45	30	30	75	45
21	22/52 rhetorical questions	45	45	30	30	45
22	23/52 linking (i)	210	150	420	180	135
23	24/52 linking (ii)	120	105	255	210	90
24	25/52 linking (iii)	75	90	45	0	75



25	26/52 identifying transitions	60	30	15	60	45
26	27/52 identifying transitions (ii)	300	194	165	45	195
27	28/52 note taking (i)	45	76	15	195	0
28	29/52 note taking (ii)	75	30	150	30	0
29	30/52 note taking (iii)		570	405	270	0
30	31/52 note taking (iv)			30	15	0
31	32/52 note taking (v)			120	60	0
32	33/52 note taking (vi)			15	30	0
33	34/52 note taking (vii)			30	15	0
34	35/52 note taking (viii)			15		0
35	36/52 note taking (ix)			105		0
36	37/52 Note taking (x)			15		0
37	38/52 Idioms(i)			15		15
38	39 /52 idioms(ii)			0		30
39	40/52 Academic English (i)			75		
40	41/52 academic English (ii)			15		
41	42/52 Academic English (iii)			15		
42						
43						
44						
45						
46						
47						
48						
49						
50						
51						
52						
	Time in Seconds	2835	2760	3150	2895	1515
	Time in Mins	47.25	46	52.5	48.25	25.25

APPENDIX 7.5 CD CONTAINING ADDITIONAL MATERIALS:VISIO FLOWCHARTS

1. Flowcharts of 20 students' navigation
2. Summary of navigation in narrative
3. Screenshots from activity monitor of a few students
4. Key logger files of few students
5. Visio flowcharts converted to GIF